

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization  
International Bureau(43) International Publication Date  
10 April 2003 (10.04.2003)

PCT

(10) International Publication Number  
**WO 03/029459 A2**(51) International Patent Classification<sup>7</sup>: **C12N 15/11,**  
A61K 48/0037083 Göttingen (DE). RAUHUT, Reinhard [DE/DE];  
Merkelstrasse 16, 37085 Göttingen (DE).

(21) International Application Number: PCT/EP02/10881

(74) Agents: WEICKMANN, Franz, Albert et al.; Weick-  
mann & Weickmann, Postfach 860 820, 81635 München  
(DE).

(22) International Filing Date:

27 September 2002 (27.09.2002)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

01123453.1	28 September 2001 (28.09.2001)	EP
02006712.0	22 March 2002 (22.03.2002)	EP
02016772.2	26 July 2002 (26.07.2002)	EP

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU,  
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU,  
CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH,  
GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC,  
LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW,  
MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG,  
SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ,  
VC, VN, YU, ZA, ZM, ZW.(84) Designated States (*regional*): ARIPO patent (GH, GM,  
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW),  
Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),  
European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE,  
ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK,  
TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ,  
GW, ML, MR, NE, SN, TD, TG).(71) Applicant (*for all designated States except US*): MAX-  
PLANCK-GESELLSCHAFT ZUR FÖRDERUNG  
DER WISSENSCHAFTEN E.V. [DE/DE]; Hofgarten-  
strasse 8, 80539 München (DE).

(72) Inventors; and

(75) Inventors/Applicants (*for US only*): TUSCHL, Thomas  
[DE/DE]; Keplerstrasse 9, 37085 Göttingen (DE).  
LAGOS-QUINTANA, Mariana [MX/DE]; Kurze Geis-  
marstrasse 13, 37073 Göttingen (DE). LENDECKEL,  
Winfried [DE/DE]; Blinde Gasse 52, 37318 Hohengand-  
ern (DE). MEYER, Jutta [DE/DE]; Sternstrasse 16,

Published:

— without international search report and to be republished  
upon receipt of that reportFor two-letter codes and other abbreviations, refer to the "Guid-  
ance Notes on Codes and Abbreviations" appearing at the begin-  
ning of each regular issue of the PCT Gazette.

(54) Title: MICRORNA MOLECULES

(57) Abstract: In *Caenorhabditis elegans*, lin-4 and let-7 encode 22- and 21 -nucleotide RNAs, respectively, that function as key regulators of developmental timing. Because the appearance of these short RNAs is regulated during development, they are also referred to as "small temporal RNAs" (stRNAs). We show that many more 21- and 22-nt expressed RNAs, termed microRNAs, (miRNAs), exist in invertebrates and vertebrates, and that some of these novel RNAs, similar to let-7 stRNA, are also highly conserved. This suggests that sequence-specific post-transcriptional regulatory mechanisms mediated by small RNAs are more general than previously appreciated.

- 1 -

**MicroRNA molecules****Description**

5

The present invention relates to novel small expressed (micro)RNA molecules associated with physiological regulatory mechanisms, particularly in developmental control.

10 In *Caenorhabditis elegans*, *lin-4* and *let-7* encode 22- and 21-nucleotide RNAs, respectively (1, 2), that function as key regulators of developmental timing (3-5). Because the appearance of these short RNAs is regulated during development, they are also referred to as "microRNAs" (miRNAs) or small temporal RNAs (stRNAs) (6). *lin-4* and *let-21* are the only known  
15 miRNAs to date.

Two distinct pathways exist in animals and plants in which 21- to 23-nucleotide RNAs function as post-transcriptional regulators of gene expression. Small interfering RNAs (siRNAs) act as mediators of sequence-specific mRNA degradation in RNA interference (RNAi) (7-11) whereas  
20 miRNAs regulate developmental timing by mediating sequence-specific repression of mRNA translation (3-5). siRNAs and miRNAs are excised from double-stranded RNA (dsRNA) precursors by Dicer (12, 13, 29), a multidomain RNase III protein, thus producing RNA species of similar size.  
25 However, siRNAs are believed to be double-stranded (8, 11, 12), while miRNAs are single-stranded (6).

We show that many more short, particularly 21- and 22-nt expressed RNAs, termed microRNAs (miRNAs), exist in invertebrates and vertebrates,  
30 and that some of these novel RNAs, similar to *let-7* RNA (6), are also highly conserved. This suggests that sequence-specific post-transcriptional

- 2 -

regulatory mechanisms mediated by small RNAs are more general than previously appreciated.

The present invention relates to an isolated nucleic acid molecule comprising:

- (a) a nucleotide sequence as shown in Table 1, Table 2, Table 3 or Table 4
- (b) a nucleotide sequence which is the complement of (a),
- (c) a nucleotide sequence which has an identity of at least 80%, preferably of at least 90% and more preferably of at least 99%, to a sequence of (a) or (b) and/or
- (d) a nucleotide sequence which hybridizes under stringent conditions to a sequence of (a), (b) and/or (c).

In a preferred embodiment the invention relates to miRNA molecules and analogs thereof, to miRNA precursor molecules and to DNA molecules encoding miRNA or miRNA precursor molecules.

Preferably the identity of sequence (c) to a sequence of (a) or (b) is at least 90%, more preferably at least 95%. The determination of identity (percent) may be carried out as follows:

$$I = n : L$$

wherein I is the identity in percent, n is the number of identical nucleotides between a given sequence and a comparative sequence as shown in Table 1, Table 2, Table 3 or Table 4 and L is the length of the comparative sequence. It should be noted that the nucleotides A, C, G and U as depicted in Tables 1, 2, 3 and 4 may denote ribonucleotides,

- 3 -

deoxyribonucleotides and/or other nucleotide analogs, e.g. synthetic non-naturally occurring nucleotide analogs. Further nucleobases may be substituted by corresponding nucleobases capable of forming analogous H-bonds to a complementary nucleic acid sequence, e.g. U may be substituted by T.

Further, the invention encompasses nucleotide sequences which hybridize under stringent conditions with the nucleotide sequence as shown in Table 1, Table 2, Table 3 or Table 4, a complementary sequence thereof or a highly identical sequence. Stringent hybridization conditions comprise washing for 1 h in 1 x SSC and 0.1% SDS at 45°C, preferably at 48°C and more preferably at 50°C, particularly for 1 h in 0.2 x SSC and 0.1% SDS.

The isolated nucleic acid molecules of the invention preferably have a length of from 18 to 100 nucleotides, and more preferably from 18 to 80 nucleotides. It should be noted that mature miRNAs usually have a length of 19-24 nucleotides, particularly 21, 22 or 23 nucleotides. The miRNAs, however, may be also provided as a precursor which usually has a length of 50-90 nucleotides, particularly 60-80 nucleotides. It should be noted that the precursor may be produced by processing of a primary transcript which may have a length of > 100 nucleotides.

The nucleic acid molecules may be present in single-stranded or double-stranded form. The miRNA as such is usually a single-stranded molecule, while the mi-precursor is usually an at least partially self-complementary molecule capable of forming double-stranded portions, e.g. stem- and loop-structures. DNA molecules encoding the miRNA and miRNA precursor molecules. The nucleic acids may be selected from RNA, DNA or nucleic acid analog molecules, such as sugar- or backbone-modified ribonucleotides or deoxyribonucleotides. It should be noted, however, that other nucleic analogs, such as peptide nucleic acids (PNA) or locked nucleic acids (LNA), are also suitable.

- 4 -

In an embodiment of the invention the nucleic acid molecule is an RNA- or DNA molecule, which contains at least one modified nucleotide analog, i.e. a naturally occurring ribonucleotide or deoxyribonucleotide is substituted by a non-naturally occurring nucleotide. The modified nucleotide analog may be located for example at the 5'-end and/or the 3'-end of the nucleic acid molecule.

Preferred nucleotide analogs are selected from sugar- or backbone-modified ribonucleotides. It should be noted, however, that also nucleobase-modified ribonucleotides, i.e. ribonucleotides, containing a non-naturally occurring nucleobase instead of a naturally occurring nucleobase such as uridines or cytidines modified at the 5-position, e.g. 5-(2-amino)propyl uridine, 5-bromo uridine; adenosines and guanosines modified at the 8-position, e.g. 8-bromo guanosine; deaza nucleotides, e.g. 7-deaza-adenosine; O- and N-alkylated nucleotides, e.g. N6-methyl adenosine are suitable. In preferred sugar-modified ribonucleotides the 2'-OH-group is replaced by a group selected from H, OR, R, halo, SH, SR, NH<sub>2</sub>, NHR, NR<sub>2</sub> or CN, wherein R is C<sub>1</sub>-C<sub>8</sub> alkyl, alkenyl or alkynyl and halo is F, Cl, Br or I. In preferred backbone-modified ribonucleotides the phosphoester group connecting to adjacent ribonucleotides is replaced by a modified group, e.g. of phosphothioate group. It should be noted that the above modifications may be combined.

The nucleic acid molecules of the invention may be obtained by chemical synthesis methods or by recombinant methods, e.g. by enzymatic transcription from synthetic DNA-templates or from DNA-plasmids isolated from recombinant organisms. Typically phage RNA-polymerases are used for transcription, such as T7, T3 or SP6 RNA-polymerases.

The invention also relates to a recombinant expression vector comprising a recombinant nucleic acid operatively linked to an expression control sequence, wherein expression, i.e. transcription and optionally further

- 5 -

processing results in a miRNA-molecule or miRNA precursor molecule as described above. The vector is preferably a DNA-vector, e.g. a viral vector or a plasmid, particularly an expression vector suitable for nucleic acid expression in eukaryotic, more particularly mammalian cells. The recombinant nucleic acid contained in said vector may be a sequence  
5 which results in the transcription of the miRNA-molecule as such, a precursor or a primary transcript thereof, which may be further processed to give the miRNA-molecule.

10 Further, the invention relates to diagnostic or therapeutic applications of the claimed nucleic acid molecules. For example, miRNAs may be detected in biological samples, e.g. in tissue sections, in order to determine and classify certain cell types or tissue types or miRNA-associated pathogenic disorders which are characterized by differential expression of miRNA-  
15 molecules or miRNA-molecule patterns. Further, the developmental stage of cells may be classified by determining temporarily expressed miRNA-molecules.

Further, the claimed nucleic acid molecules are suitable for therapeutic  
20 applications. For example, the nucleic acid molecules may be used as modulators or targets of developmental processes or disorders associated with developmental dysfunctions, such as cancer. For example, miR-15 and miR-16 probably function as tumor-suppressors and thus expression or delivery of these RNAs or analogs or precursors thereof to tumor cells may  
25 provide therapeutic efficacy, particularly against leukemias, such as B-cell chronic lymphocytic leukemia (B-CLL). Further, miR-10 is a possible regulator of the translation of Hox Genes, particularly Hox 3 and Hox 4 (or Scr and Dfd in Drosophila).

30 In general, the claimed nucleic acid molecules may be used as a modulator of the expression of genes which are at least partially complementary to said nucleic acid. Further, miRNA molecules may act as target for

- 6 -

therapeutic screening procedures, e.g. inhibition or activation of miRNA molecules might modulate a cellular differentiation process, e.g. apoptosis.

Furthermore, existing miRNA molecules may be used as starting materials  
5 for the manufacture of sequence-modified miRNA molecules, in order to modify the target-specificity thereof, e.g. an oncogene, a multidrug-resistance gene or another therapeutic target gene. The novel engineered miRNA molecules preferably have an identity of at least 80% to the starting miRNA, e.g. as depicted in Tables 1, 2, 3 and 4. Further, miRNA  
10 molecules can be modified, in order that they are symmetrically processed and then generated as double-stranded siRNAs which are again directed against therapeutically relevant targets.

Furthermore, miRNA molecules may be used for tissue reprogramming  
15 procedures, e.g. a differentiated cell line might be transformed by expression of miRNA molecules into a different cell type or a stem cell.

For diagnostic or therapeutic applications, the claimed RNA molecules are preferably provided as a pharmaceutical composition. This pharmaceutical  
20 composition comprises as an active agent at least one nucleic acid molecule as described above and optionally a pharmaceutically acceptable carrier.

The administration of the pharmaceutical composition may be carried out  
25 by known methods, wherein a nucleic acid is introduced into a desired target cell in vitro or in vivo.

Commonly used gene transfer techniques include calcium phosphate, DEAE-dextran, electroporation and microinjection and viral methods [30,  
30 31, 32, 33, 34]. A recent addition to this arsenal of techniques for the introduction of DNA into cells is the use of cationic liposomes [35].

- 7 -

Commercially available cationic lipid formulations are e.g. Tfx 50 (Promega) or Lipofectamin 2000 (Life Technologies).

5 The composition may be in form of a solution, e.g. an injectable solution, a cream, ointment, tablet, suspension or the like. The composition may be administered in any suitable way, e.g. by injection, by oral, topical, nasal, rectal application etc. The carrier may be any suitable pharmaceutical carrier. Preferably, a carrier is used, which is capable of increasing the efficacy of the RNA molecules to enter the target-cells. Suitable examples  
10 of such carriers are liposomes, particularly cationic liposomes.

Further, the invention relates to a method of identifying novel microRNA-molecules and precursors thereof, in eukaryotes, particularly in vertebrates and more particularly in mammals, such as humans or mice. This method  
15 comprises: ligating 5'- and 3'-adapter-molecules to the end of a size-fractionated RNA-population, reverse transcribing said adapter-ligated RNA-population, and characterizing said reverse transcribed RNA-molecules, e.g. by amplification, concatamerization, cloning and sequencing.

20 A method as described above already has been described in (8), however, for the identification of siRNA molecules. Surprisingly, it was found now that the method is also suitable for identifying the miRNA molecules or precursors thereof as claimed in the present application.

25 Further, it should be noted that as 3'-adaptor for derivatization of the 3'-OH group not only 4-hydroxymethylbenzyl but other types of derivatization groups, such as alkyl, alkyl amino, ethylene glycol or 3'-deoxy groups are suitable.

30 Further, the invention shall be explained in more detail by the following Figures and Examples:



## Figure Legends

Fig. 1A. Expression of *D. melanogaster* miRNAs. Northern blots of total RNA isolated from staged populations of *D. melanogaster* were probed for the indicated miRNAs. The position of 76-nt val-tRNA is also indicated on the blots. 5S rRNA serves as loading control. E, embryo; L, larval stage; P, pupae; A, adult; S2, Schneider-2 cells. It should be pointed out, that S2 cells are polyclonal, derived from an unknown subset of embryonic tissues, and may have also lost some features of their tissue of origin while maintained in culture. miR-3 to miR-6 RNAs were not detectable in S2 cells (data not shown). miR-14 was not detected by Northern blotting and may be very weakly expressed, which is consistent with its cloning frequency. Similar miRNA sequences are difficult to distinguish by Northern blotting because of potential cross-hybridization of probes.

Fig. 1B. Expression of vertebrate miRNAs. Northern blots of total RNA isolated from HeLa cells, mouse kidneys, adult zebrafish, frog ovaries, and S2 cells were probed for the indicated miRNAs. The position of 76-nt val-tRNA is also indicated on the blots. 5S rRNA from the preparations of total RNA from the indicated species is also shown. The gels used for probing of miR-18, miR-19a, miR-30, and miR-31 were not run as far as the other gels (see tRNA marker position). miR-32 and miR-33 were not detected by Northern blotting, which is consistent with their low cloning frequency. Oligodeoxynucleotides used as Northern probes were:

let-7a, 5' TACTATACAACCTACTACCTCAATTTGCC (SEQ ID NO:1);  
 let-7d, 5' ACTATGCAACCTACTACCTCT (SEQ ID NO:2);  
 let-7e, 5' ACTATACAACCTCCTACCTCA (SEQ ID NO:3);  
*D. melanogaster* val-tRNA, 5' TGGTGTTTCCGCCCGGGAA (SEQ ID NO:4);  
 miR-1, 5' TGGAATGTAAAGAAGTATGGAG (SEQ ID NO:5);  
 miR-2b, 5' GCTCCTCAAAGCTGGCTGTGATA (SEQ ID NO:6);  
 miR-3, 5' TGAGACACACTTTGCCAGTGA (SEQ ID NO:7);  
 miR-4, 5' TCAATGGTTGTCTAGCTTTAT (SEQ ID NO:8);

- 9 -

miR-5, 5' CATATCACAACGATCGTTCCTTT (SEQ ID NO:9);  
 miR-6, 5' AAAAAGAACAGCCACTGTGATA (SEQ ID NO:10);  
 miR-7, 5' TGGAAGACTAGTGATTTTGTGTTGT (SEQ ID NO:11);  
 miR-8, 5' GACATCTTTACCTGACAGTATTA (SEQ ID NO:12);  
 5 miR-9, 5' TCATACAGCTAGATAACCAAAGA (SEQ ID NO:13);  
 miR-10, 5' ACAAATTCGGATCTACAGGGT (SEQ ID NO:14);  
 miR-11, 5' GCAAGAACTCAGACTGTGATG (SEQ ID NO:15);  
 miR-12, 5' ACCAGTACCTGATGTAATACTCA (SEQ ID NO:16);  
 miR-13a, 5' ACTCGTCAAAATGGCTGTGATA (SEQ ID NO:17);  
 10 miR-14, 5' TAGGAGAGAGAAAAAGACTGA (SEQ ID NO:18);  
 miR-15, 5' TAGCAGCACATAATGGTTTGT (SEQ ID NO:19);  
 miR-16, 5' GCCAATATTTACGTGCTGCTA (SEQ ID NO:20);  
 miR-17, 5' TACAAGTGCCTTCACTGCAGTA (SEQ ID NO:21);  
 miR-18, 5' TATCTGCACTAGATGCACCTTA (SEQ ID NO:22);  
 15 miR-19a, 5' TCAGTTTTGCATAGATTTGCACA (SEQ ID NO:23);  
 miR-20, 5' TACCTGCACTATAAGCACTTTA (SEQ ID NO:24);  
 miR-21, 5' TCAACATCAGTCTGATAAGCTA (SEQ ID NO:25);  
 miR-22, 5' ACAGTTCTTCAACTGGCAGCTT (SEQ ID NO:26);  
 miR-23, 5' GGAAATCCCTGGCAATGTGAT (SEQ ID NO:27);  
 20 miR-24, 5' CTGTTCTGCTGAACTGAGCCA (SEQ ID NO:28);  
 miR-25, 5' TCAGACCGAGACAAGTGCAATG (SEQ ID NO:29);  
 miR-26a, 5' AGCCTATCCTGGATTACTTGAA (SEQ ID NO:30);  
 miR-27, 5' AGCGGAACTTAGCCACTGTGAA (SEQ ID NO:31);  
 miR-28, 5' CTCAATAGACTGTGAGCTCCTT (SEQ ID NO:32);  
 25 miR-29, 5' AACCGATTTCAGATGGTGCTAG (SEQ ID NO:33);  
 miR-30, 5' GCTGCAAACATCCGACTGAAAG (SEQ ID NO:34);  
 miR-31, 5' CAGCTATGCCAGCATCTTGCCT (SEQ ID NO:35);  
 miR-32, 5' GCAACTTAGTAATGTGCAATA (SEQ ID NO:36);  
 miR-33, 5' TGCAATGCAACTACAATGCACC (SEQ ID NO:37).

30

Fig. 2. Genomic organization of miRNA gene clusters. The precursor structure is indicated as box and the location of the miRNA within the

- 10 -

precursor is shown in gray; the chromosomal location is also indicated to the right. (A) *D. melanogaster* miRNA gene clusters. (B) Human miRNA gene clusters. The cluster of let-7a-1 and let-7f-1 is separated by 26500 nt from a copy of let-7d on chromosome 9 and 17. A cluster of let-7a-3 and let-7b, separated by 938 nt on chromosome 22, is not illustrated.

Fig. 3. Predicted precursor structures of *D. melanogaster* miRNAs. RNA secondary structure prediction was performed using mfold version 3.1 [28] and manually refined to accommodate G/U wobble base pairs in the helical segments. The miRNA sequence is underlined. The actual size of the stem-loop structure is not known experimentally and may be slightly shorter or longer than represented. Multicopy miRNAs and their corresponding precursor structures are also shown.

Fig. 4. Predicted precursor structures of human miRNAs. For legend, see Fig. 3.

Fig. 5. Expression of novel mouse miRNAs. Northern blot analysis of novel mouse miRNAs. Total RNA from different mouse tissues was blotted and probed with a 5'-radiolabeled oligodeoxynucleotide complementary to the indicated miRNA. Equal loading of total RNA on the gel was verified by ethidium bromide staining prior to transfer; the band representing tRNAs is shown. The fold-back precursors are indicated with capital L. Mouse brains were dissected into midbrain, mb, cortex, cx, cerebellum, cb. The rest of the brain, rb, was also used. Other tissues were heart, ht, lung, lg, liver, lv, colon, co, small intestine, si, pancreas, pc, spleen, sp, kidney, kd, skeletal muscle, sm, stomach, st, H, human Hela SS3 cells. Oligodeoxynucleotides used as Northern probes were:

miR-1a, CTCCATACTTCTTTACATTCCA (SEQ ID NO:38);

miR-30b, GCTGAGTGTAGGATGTTTACA (SEQ ID NO:39);

miR-30a-s, GCTTCCAGTCGAGGATGTTTACA (SEQ ID NO:40);

miR-99b, CGCAAGGTCGGTTCTACGGGTG (SEQ ID NO:41);

- 11 -

miR-101, TCAGTTATCACAGTACTGTA (SEQ ID NO:42);  
 miR-122a, ACAAACACCATTTGTCACACTCCA (SEQ ID NO:43);  
 miR-124a, TGGCATTACACGCGTGCCTTA (SEQ ID NO:44);  
 miR-125a, CACAGGTTAAAGGGTCTCAGGGA (SEQ ID NO:45);  
 5 miR-125b, TCACAAGTTAGGGTCTCAGGGA (SEQ ID NO:46);  
 miR-127, AGCCAAGCTCAGACGGATCCGA (SEQ ID NO:47);  
 miR-128, AAAAGAGACCGGTTCACTCTGA (SEQ ID NO:48);  
 miR-129, GCAAGCCCAGACCGAAAAAAG (SEQ ID NO:49);  
 miR-130, GCCCTTTTAACATTGCACTC (SEQ ID NO:50);  
 10 miR-131, ACTTTCGGTTATCTAGCTTTA (SEQ ID NO:51);  
 miR-132, ACGACCATGGCTGTAGACTGTTA (SEQ ID NO:52);  
 miR-143, TGAGCTACAGTGCTTCATCTCA (SEQ ID NO:53).

15 Fig.6. Potential orthologs of lin-4 stRNA. (A) Sequence alignment of *C. elegans* lin-4 stRNA with mouse miR-125a and miR-125b and the *D. melanogaster* miR-125. Differences are highlighted by gray boxes. (B) Northern blot of total RNA isolated from staged populations of *D. melanogaster*, probed for miR-125. E, embryo; L, larval stage; P, pupae; A, adult; S2, Schneider-2 cells.  
 20

Fig. 7. Predicted precursor structures of miRNAs, sequence accession numbers and homology information. RNA secondary structure prediction was performed using mfold version 3.1 and manually refined to  
 25 accommodate G/U wobble base pairs in the helical segments. Dashes were inserted into the secondary structure presentation when asymmetrically bulged nucleotides had to be accommodated. The excised miRNA sequence is underlined. The actual size of the stem-loop structure is not known experimentally and may be slightly shorter or longer than  
 30 represented. Multicopy miRNAs and their corresponding precursor structures are also shown. In cases where no mouse precursors were yet deposited in the database, the human orthologs are indicated. miRNAs

- 12 -

which correspond to *D. melanogaster* or human sequences are included. Published *C. elegans* miRNAs [36, 37] are also included in the table. A recent set of new HeLa cell miRNAs is also indicated [46]. If several ESTs were retrieved for one organism in the database, only those with different precursor sequences are listed. miRNA homologs found in other species are indicated. Chromosomal location and sequence accession numbers, and clusters of miRNA genes are indicated. Sequences from cloned miRNAs were searched against mouse and human in GenBank (including trace data), and against *Fugu rubripes* and *Danio rerio* at [www.jgi.doe.gov](http://www.jgi.doe.gov) and [www.sanger.ac.uk](http://www.sanger.ac.uk), respectively.

EXAMPLE 1: MicroRNAs from *D. melanogaster* and human.

We previously developed a directional cloning procedure to isolate siRNAs after processing of long dsRNAs in *Drosophila melanogaster* embryo lysate (8). Briefly, 5' and 3' adapter molecules were ligated to the ends of a size-fractionated RNA population, followed by reverse transcription, PCR amplification, concatamerization, cloning and sequencing. This method, originally intended to isolate siRNAs, led to the simultaneous identification of 14 novel 20- to 23-nt short RNAs which are encoded in the *D. melanogaster* genome and which are expressed in 0 to 2 h embryos (Table 1). The method was adapted to clone RNAs in a similar size range from HeLa cell total RNA (14), which led to the identification of 19 novel human stRNAs (Table 2), thus providing further evidence for the existence of a large class of small RNAs with potential regulatory roles. According to their small size, we refer to these novel RNAs as microRNAs or miRNAs. The miRNAs are abbreviated as miR-1 to miR-33, and the genes encoding miRNAs are named mir-1 to mir-33. Highly homologous miRNAs are classified by adding a lowercase letter, followed by a dash and a number for designating multiple genomic copies of a mir gene.

- 13 -

The expression and size of the cloned, endogenous short RNAs was also examined by Northern blotting (Fig. 1, Table 1 and 2). Total RNA isolation was performed by acid guanidinium thiocyanate-phenol-chloroform extraction [45]. Northern analysis was performed as described [1], except that the total RNA was resolved on a 15% denaturing polyacrylamide gel, transferred onto Hybond-N+ membrane (Amersham Pharmacia Biotech), and the hybridization and wash steps were performed at 50°C. Oligodeoxynucleotides used as Northern probes were 5'-32P-phosphorylated, complementary to the miRNA sequence and 20 to 25 nt in length.

5S rRNA was detected by ethidium staining of polyacrylamide gels prior to transfer. Blots were stripped by boiling in 0.1% aqueous sodium dodecylsulfate/0.1x SSC (15 mM sodium chloride, 1.5 mM sodium citrate, pH 7.0) for 10 min, and were re-probed up to 4 times until the 21-nt signals became too weak for detection. Finally, blots were probed for val-tRNA as size marker.

For analysis of *D. melanogaster* RNAs, total RNA was prepared from different developmental stages, as well as cultured Schneider-2 (S2) cells, which originally derive from 20-24 h *D. melanogaster* embryos [15] (Fig. 1, Table 1). miR-3 to miR-7 are expressed only during embryogenesis and not at later developmental stages. The temporal expression of miR-1, miR-2 and miR-8 to miR-13 was less restricted. These miRNAs were observed at all developmental stages though significant variations in the expression levels were sometimes observed. Interestingly, miR-1, miR-3 to miR-6, and miR-8 to miR-11 were completely absent from cultured Schneider-2 (S2) cells, which were originally derived from 20-24 h *D. melanogaster* embryos [15], while miR-2, miR-7, miR-12, and miR-13 were present in S2 cells, therefore indicating cell type-specific miRNA expression. miR-1, miR-8, and miR-12 expression patterns are similar to those of *lin-4* stRNA in *C. elegans*, as their expression is strongly upregulated in larvae and sustained

- 14 -

to adulthood [16]. miR-9 and miR-11 are present at all stages but are strongly reduced in the adult which may reflect a maternal contribution from germ cells or expression in one sex only.

5 The mir-3 to mir-6 genes are clustered (Fig. 2A), and mir-6 is present as triple repeat with slight variations in the mir-6 precursor sequence but not in the miRNA sequence itself. The expression profiles of miR-3 to miR-6 are highly similar (Table 1), which suggests that a single embryo-specific precursor transcript may give rise to the different miRNAs, or that the  
10 same enhancer regulates miRNA-specific promoters. Several other fly miRNAs are also found in gene clusters (Fig. 2A).

The expression of HeLa cell miR-15 to miR-33 was examined by Northern blotting using HeLa cell total RNA, in addition to total RNA prepared from  
15 mouse kidneys, adult zebrafish, *Xenopus laevis* ovary, and *D. melanogaster* S2 cells (Fig. 1B, Table 2). miR-15 and miR-16 are encoded in a gene cluster (Fig. 2B) and are detected in mouse kidney, fish, and very weakly in frog ovary, which may result from miRNA expression in somatic ovary tissue rather than oocytes. mir-17 to mir-20 are also clustered (Fig. 2B),  
20 and are expressed in HeLa cells and fish, but undetectable in mouse kidney and frog ovary (Fig. 1, Table 2), and therefore represent a likely case of tissue-specific miRNA expression.

The majority of vertebrate and invertebrate miRNAs identified in this study  
25 are not related by sequence, but a few exceptions, similar to the highly conserved let-7 RNA [6], do exist. Sequence analysis of the *D. melanogaster* miRNAs revealed four such examples of sequence conservation between invertebrates and vertebrates. miR-1 homologs are encoded in the genomes of *C. elegans*, *C. briggsae*, and humans, and are  
30 found in cDNAs from zebrafish, mouse, cow and human. The expression of mir-1 was detected by Northern blotting in total RNA from adult zebrafish and *C. elegans*, but not in total RNA from HeLa cells or mouse kidney

- 15 -

(Table 2 and data not shown). Interestingly, while mir-1 and let-7 are expressed both in adult flies (Fig. 1A) [6] and are both undetected in S2 cells, miR-1 is, in contrast to let-7, undetectable in HeLa cells. This represents another case of tissue-specific expression of a miRNA, and indicates that miRNAs may not only play a regulatory role in developmental timing, but also in tissue specification. miR-7 homologs were found by database searches in mouse and human genomic and expressed sequence tag sequences (ESTs). Two mammalian miR-7 variants are predicted by sequence analysis in mouse and human, and were detected by Northern blotting in HeLa cells and fish, but not in mouse kidney (Table 2). Similarly, we identified mouse and human miR-9 and miR-10 homologs by database searches but only detected mir-10 expression in mouse kidney.

The identification of evolutionary related miRNAs, which have already acquired multiple sequence mutations, was not possible by standard bioinformatic searches. Direct comparison of the *D. melanogaster* miRNAs with the human miRNAs identified an 11-nt segment shared between *D. melanogaster* miR-6 and HeLa miR-27, but no further relationships were detected. One may speculate that most miRNAs only act on a single target and therefore allow for rapid evolution by covariation, and that highly conserved miRNAs act on more than one target sequence, and therefore have a reduced probability for evolutionary drift by covariation [6]. An alternative interpretation is that the sets of miRNAs from *D. melanogaster* and humans are fairly incomplete and that many more miRNAs remain to be discovered, which will provide the missing evolutionary links.

lin-4 and let-7 stRNAs were predicted to be excised from longer transcripts that contain approximately 30 base-pair stem-loop structures [1, 6]. Database searches for newly identified miRNAs revealed that all miRNAs are flanked by sequences that have the potential to form stable stem-loop structures (Fig. 3 and 4). In many cases, we were able to detect the predicted, approximately 70-nt precursors by Northern blotting (Fig. 1).



- 16 -

Some miRNA precursor sequences were also identified in mammalian cDNA (EST) databases [27], indicating that primary transcripts longer than 70-nt stem-loop precursors do also exist. We never cloned a 22-nt RNA complementary to any of the newly identified miRNAs, and it is as yet  
5 unknown how the cellular processing machinery distinguishes between the miRNA and its complementary strand. Comparative analysis of the precursor stem-loop structures indicates that the loops adjacent to the base-paired miRNA segment can be located on either side of the miRNA sequence (Fig. 3 and 4), suggesting that the 5' or 3' location of the stem-  
10 closing loop is not the determinant of miRNA excision. It is also unlikely that the structure, length or stability of the precursor stem is the critical determinant as the base-paired structures are frequently imperfect and interspersed by less stable, non-Watson-Crick base pairs such as G/A, U/U, C/U, A/A, and G/U wobbles. Therefore, a sequence-specific recognition  
15 process is a likely determinant for miRNA excision, perhaps mediated by members of the Argonaute (rde-1/ago1/piwi) protein family. Two members of this family, alg-1 and alg-2, have recently been shown to be critical for stRNA processing in *C. elegans* [13]. Members of the Argonaute protein family are also involved in RNAi and PTGS. In *D. melanogaster*, these  
20 include argonaute2, a component of the siRNA-endonuclease complex (RISC) [17], and its relative aubergine, which is important for silencing of repeat genes [18]. In other species, these include rde-1, argonaute1, and qde-2, in *C. elegans* [19], *Arabidopsis thaliana* [20], and *Neurospora crassa* [21], respectively. The Argonaute protein family therefore represents,  
25 besides the RNase III Dicer [12, 13], another evolutionary link between RNAi and miRNA maturation.

Despite advanced genome projects, computer-assisted detection of genes encoding functional RNAs remains problematic [22]. Cloning of expressed,  
30 short functional RNAs, similar to EST approaches (RNomics), is a powerful alternative and probably the most efficient method for identification of such novel gene products [23-26]. The number of functional RNAs has been

- 17 -

widely underestimated and is expected to grow rapidly because of the development of new functional RNA cloning methodologies.

The challenge for the future is to define the function and the potential  
5 targets of these novel miRNAs by using bioinformatics as well as genetics,  
and to establish a complete catalogue of time- and tissue-specific  
distribution of the already identified and yet to be uncovered miRNAs. lin-4  
and let-7 stRNAs negatively regulate the expression of proteins encoded by  
mRNAs whose 3' untranslated regions contain sites of complementarity to  
10 the stRNA [3-5].

Thus, a series of 33 novel genes, coding for 19- to 23-nucleotide  
microRNAs (miRNAs), has been cloned from fly embryos and human cells.  
Some of these miRNAs are highly conserved between vertebrates and  
15 invertebrates and are developmentally or tissue-specifically expressed. Two  
of the characterized human miRNAs may function as tumor suppressors in  
B-cell chronic lymphocytic leukemia. miRNAs are related to a small class of  
previously described 21- and 22-nt RNAs (lin-4 and let-7 RNAs), so-called  
small temporal RNAs (stRNAs), and regulate developmental timing in C.  
20 elegans and other species. Similar to stRNAs, miRNAs are presumed to  
regulate translation of specific target mRNAs by binding to partially  
complementary sites, which are present in their 3'-untranslated regions.

Deregulation of miRNA expression may be a cause of human disease, and  
25 detection of expression of miRNAs may become useful as a diagnostic.  
Regulated expression of miRNAs in cells or tissue devoid of particular  
miRNAs may be useful for tissue engineering, and delivery or transgenic  
expression of miRNAs may be useful for therapeutic intervention. miRNAs  
may also represent valuable drug targets itself. Finally, miRNAs and their  
30 precursor sequences may be engineered to recognize therapeutic valuable  
targets.

- 18 -

EXAMPLE 2: miRNAs from mouse.

To gain more detailed insights into the distribution and function of miRNAs in mammals, we investigated the tissue-specific distribution of miRNAs in adult mouse. Cloning of miRNAs from specific tissues was preferred over whole organism-based cloning because low-abundance miRNAs that normally go undetected by Northern blot analysis are identified clonally. Also, in situ hybridization techniques for detecting 21-nt RNAs have not yet been developed. Therefore, 19- to 25-nucleotide RNAs were cloned and sequenced from total RNA, which was isolated from 18.5-weeks old BL6 mice. Cloning of miRNAs was performed as follows: 0.2 to 1 mg of total RNA was separated on a 15% denaturing polyacrylamide gel and RNA of 19- to 25-nt size was recovered. A 5'-phosphorylated 3'-adapter oligonucleotide (5'-pUUUaaccgcgaattccagx: uppercase, RNA; lowercase, DNA; p, phosphate; x, 3'-Amino-Modifier C-7, ChemGenes, Ashland, Ma, USA, Cat. No. NSS-1004; SEQ ID NO:54) and a 5'-adapter oligonucleotide (5'-acggaattcctcactAAA: uppercase, RNA; lowercase, DNA; SEQ ID NO:55) were ligated to the short RNAs. RT/PCR was performed with 3'-primer (5'-GACTAGCTGGAATTCGCGGTTAAA; SEQ ID NO:56) and 5'-primer (5'-CAGCCAACGGAATTCCTCACTAAA; SEQ ID NO:57). In order to introduce Ban I restriction sites, a second PCR was performed using the primer pair 5'-CAGCCAACAGGCACCGAATTCCTCACTAAA (SEQ ID NO:57) and 5'-GACTAGCTTGGTGCCGAATTCGCGGTTAAA (SEQ ID NO:56), followed by concatamerization after Ban I digestion and T4 DNA ligation. Concatamers of 400 to 600 basepairs were cut out from 1.5% agarose gels and recovered by Biotrap (Schleicher & Schuell) electroelution (1x TAE buffer) and by ethanol precipitation. Subsequently, the 3' ends of the concatamers were filled in by incubating for 15 min at 72°C with Taq polymerase in standard PCR reaction mixture. This solution was diluted 3-fold with water and directly used for ligation into pCR2.1 TOPO vectors. Clones were screened for inserts by PCR and 30 to 50 samples were subjected to sequencing. Because RNA was prepared from combining

- 19 -

tissues of several mice, minor sequence variations that were detected multiple times in multiple clones may reflect polymorphisms rather than RT/PCR mutations. Public database searching was used to identify the genomic sequences encoding the approx. 21-nt RNAs. The occurrence of  
5 a 20 to 30 basepair fold-back structure involving the immediate upstream or downstream flanking sequences was used to assign miRNAs [36-38].

We examined 9 different mouse tissues and identified 34 novel miRNAs, some of which are highly tissue-specifically expressed (Table 3 and Figure  
10 5). Furthermore, we identified 33 new miRNAs from different mouse tissues and also from human Soas-2 osteosarcoma cells (Table 4). miR-1 was previously shown by Northern analysis to be strongly expressed in adult heart, but not in brain, liver, kidney, lung or colon [37]. Here we show that miR-1 accounts for 45% of all mouse miRNAs found in heart,  
15 yet miR-1 was still expressed at a low level in liver and midbrain even though it remained undetectable by Northern analysis. Three copies or polymorphic alleles of miR-1 were found in mice. The conservation of tissue-specific miR-1 expression between mouse and human provides additional evidence for a conserved regulatory role of this miRNA. In liver,  
20 variants of miR-122 account for 72% of all cloned miRNAs and miR-122 was undetected in all other tissues analyzed. In spleen, miR-143 appeared to be most abundant, at a frequency of approx. 30%. In colon, miR-142-as, was cloned several times and also appeared at a frequency of 30%. In small intestine, too few miRNA sequences were obtained to permit  
25 statistical analysis. This was due to strong RNase activity in this tissue, which caused significant breakdown of abundant non-coding RNAs, e.g. rRNA, so that the fraction of miRNA in the cloned sequences was very low. For the same reason, no miRNA sequences were obtained from  
pancreas.

30 To gain insights in neural tissue miRNA distribution, we analyzed cortex, cerebellum and midbrain. Similar to heart, liver and small intestine, variants

- 20 -

of a particular miRNA, miR-124, dominated and accounted for 25 to 48% of all brain miRNAs. miR-101, -127, -128, -131, and -132, also cloned from brain tissues, were further analyzed by Northern blotting and shown to be predominantly brain-specific. Northern blot analysis was performed as described in Example 1. tRNAs and 5S rRNA were detected by ethidium staining of polyacrylamide gels prior to transfer to verify equal loading. Blots were stripped by boiling in deionized water for 5 min, and reprobed up to 4 times until the 21-nt signals became too weak for detection.

miR-125a and miR-125b are very similar to the sequence of *C. elegans* lin-4 stRNA and may represent its orthologs (Fig. 6A). This is of great interest because, unlike let-7 that was readily detected in other species, lin-4 has acquired a few mutations in the central region and thus escaped bioinformatic database searches. Using the mouse sequence miR-125b, we could readily identify its ortholog in the *D. melanogaster* genome. miR-125a and miR-125b differ only by a central diuridine insertion and a U to C change. miR-125b is very similar to lin-4 stRNA with the differences located only in the central region, which is presumed to be bulged out during target mRNA recognition [41]. miR-125a and miR-125b were cloned from brain tissue, but expression was also detected by Northern analysis in other tissues, consistent with the role for lin-4 in regulating neuronal remodeling by controlling lin-14 expression [43]. Unfortunately, orthologs to *C. elegans* lin-14 have not been described and miR-125 targets remain to be identified in *D. melanogaster* or mammals. Finally, miR-125b expression is also developmentally regulated and only detectable in pupae and adult but not in embryo or larvae of *D. melanogaster* (Fig. 6B).

Sequence comparison of mouse miRNAs with previously described miRNA reveals that miR-99b and miR-99a are similar to *D. melanogaster*, mouse and human miR-10 as well as *C. elegans* miR-51 [36], miR-141 is similar to *D. melanogaster* miR-8, miR-29b is similar to *C. elegans* miR-83, and miR-131 and miR-142-s are similar to *D. melanogaster* miR-4 and *C.*

- 21 -

*elegans* miR-79 [36]. miR-124a is conserved between invertebrates and vertebrates. In this respect it should be noted that for almost every miRNA cloned from mouse was also encoded in the human genome, and frequently detected in other vertebrates, such as the pufferfish, *Fugu rubripes*, and the zebrafish, *Danio rerio*. Sequence conservation may point to conservation in function of these miRNAs. Comprehensive information about orthologous sequences is listed in Fig. 7.

In two cases both strands of miRNA precursors were cloned (Table 3), which was previously observed once for a *C. elegans* miRNA [36]. It is thought that the most frequently cloned strand of a miRNA precursor represents the functional miRNA, which is miR-30c-s and miR-142-as, s and as indicating the 5' or 3' side of the fold-back structure, respectively.

The mir-142 gene is located on chromosome 17, but was also found at the breakpoint junction of a t(8;17) translocation, which causes an aggressive B-cell leukemia due to strong up-regulation of a translocated MYC gene [44]. The translocated MYC gene, which was also truncated at the first exon, was located only 4-nt downstream of the 3'-end of the miR-142 precursor. This suggests that translocated MYC was under the control of the upstream miR-142 promoter. Alignment of mouse and human miR-142 containing EST sequences indicate an approximately 20 nt conserved sequence element downstream of the mir-142 hairpin. This element was lost in the translocation. It is conceivable that the absence of the conserved downstream sequence element in the putative miR-142/mRNA fusion prevented the recognition of the transcript as a miRNA precursor and therefore may have caused accumulation of fusion transcripts and overexpression of MYC.

miR-155, which was cloned from colon, is excised from the known noncoding BIC RNA [47]. BIC was originally identified as a gene transcriptionally activated by promoter insertion at a common retroviral

- 22 -

integration site in B cell lymphomas induced by avian leukosis virus. Comparison of BIC cDNAs from human, mouse and chicken revealed 78% identity over 138 nucleotides [47]. The identity region covers the miR-155 fold-back precursor and a few conserved boxes downstream of the fold-back sequence. The relatively high level of expression of BIC in lymphoid organs and cells in human, mouse and chicken implies an evolutionary conserved function, but BIC RNA has also been detected at low levels in non-hematopoietic tissues [47].

Another interesting observation was that segments of perfect complementarity to miRNAs are not observed in mRNA sequences or in genomic sequences outside the miRNA inverted repeat. Although this could be fortuitous, based on the link between RNAi and miRNA processing [11, 13, 43] it may be speculated that miRNAs retain the potential to cleave perfectly complementary target RNAs. Because translational control without target degradation could provide more flexibility it may be preferred over mRNA degradation.

In summary, 63 novel miRNAs were identified from mouse and 4 novel miRNAs were identified from human Soas-2 osteosarcoma cells (Table 3 and Table 4), which are conserved in human and often also in other non-mammalian vertebrates. A few of these miRNAs appear to be extremely tissue-specific, suggesting a critical role for some miRNAs in tissue-specification and cell lineage decisions. We may have also identified the fruitfly and mammalian ortholog of *C. elegans* lin-4 stRNA. The establishment of a comprehensive list of miRNA sequences will be instrumental for bioinformatic approaches that make use of completed genomes and the power of phylogenetic comparison in order to identify miRNA-regulated target mRNAs.

## References and Notes

1. R. C. Lee, R. L. Feinbaum, V. Ambros, *Cell* 75, 843 (1993).
2. B. J. Reinhart et al., *Nature* 403, 901 (2000).
- 5 3. V. Ambros, *Curr. Opin. Genet. Dev.* 10, 428 (2000).
4. E. G. Moss, *Curr. Biol.* 10, R436 (2000).
5. F. Slack, G. Ruvkun, *Annu. Rev. Genet.* 31, 611 (1997).
6. A. E. Pasquinelli et al., *Nature* 408, 86 (2000).
7. S. M. Elbashir et al., *Nature* 411, 494 (2001).
- 10 8. S. M. Elbashir, W. Lendeckel, T. Tuschl, *Genes & Dev.* 15, 188 (2001).
9. A. J. Hamilton, D. C. Baulcombe, *Science* 286, 950 (1999).
10. S. M. Hammond, E. Bernstein, D. Beach, G. J. Hannon, *Nature* 404, 293 (2000).
- 15 11. P. D. Zamore, T. Tuschl, P. A. Sharp, D. P. Bartel, *Cell* 101, 25 (2000).
12. G. Hutvágner, J. McLachlan, É. Bálint, T. Tuschl, P. D. Zamore, *Science* 93, 834 (2001).
13. A. Grishok et al., *Cell* 106, 23 (2001).
- 20 14. Cloning of 19- to 24-nt RNAs from *D. melanogaster* 0-2 h embryo lysate was performed as described (8). For cloning of HeLa miRNAs, 1 mg of HeLa total RNA was separated on a 15% denaturing polyacrylamide gel and RNA of 19- to 25-nt size was recovered. A 5' phosphorylated 3' adapter oligonucleotide (5' pUUU-aaccgcgaattccagx: uppercase, RNA; lowercase, DNA; p, phosphate; x, 4-hydroxymethylbenzyl; SEQ ID NO:54) and a 5' adapter oligonucleotide (5' acggaattcctcactAAA: uppercase, RNA; lowercase, DNA; SEQ ID NO:55) were ligated to the short HeLa cell RNAs. RT/PCR was performed with 3' primer (5' GACTAGCTGGAATTCGCGGTTAAA; SEQ ID NO:56) and 5' primer (5' CAGCCAACGGAATTCCTCACTAAA; SEQ ID NO:57), and followed by concatamerization after Eco RI digestion and T4 DNA
- 25
- 30



- 24 -

ligation (8). After ligation of concatamers into pCR2.1 TOPO vectors, about 100 clones were selected and subjected to sequencing.

15. I. Schneider, *J Embryol Exp Morphol* 27, 353 (1972).
- 5 16. R. Feinbaum, V. Ambros, *Dev. Biol.* 210, 87 (1999).
17. S. M. Hammond, S. Boettcher, A. A. Caudy, R. Kobayashi, G. J. Hannon, *Science* 293, 1146 (2001).
18. A. A. Aravin et al., *Curr. Biol.* 11, 1017 (2001).
19. H. Tabara et al., *Cell* 99, 123 (1999).
- 10 20. M. Fagard, S. Boutet, J. B. Morel, C. Bellini, H. Vaucheret, *Proc. Natl. Acad. Sci. USA* 97, 11650 (2000).
21. C. Catalanotto, G. Azzalin, G. Macino, C. Cogoni, *Nature* 404, 245 (2000).
22. S. R. Eddy, *Curr. Opin. Genet. Dev.* 9, 695 (1999).
- 15 23. J. Cavaille et al., *Proc. Natl. Acad. Sci. USA* 97, 14311 (2000).
24. A. Hüttenhofer et al., *EMBO J.* 20, 2943 (2001).
25. L. Argaman et al., *Curr. Biol.* 11, 941 (2001).
26. K. M. Wassarman, F. Repoila, C. Rosenow, G. Storz, S. Gottesman, *Genes & Dev.* 15, 1637 (2001).
- 20 27. Supplementary Web material is available on Science Online at [www.sciencemag.org/cgi/content/full/xxx](http://www.sciencemag.org/cgi/content/full/xxx)
28. D. H. Mathews, J. Sabina, M. Zuker, D. H. Turner, *J. Mol. Biol.* 288, 911 (1999).
29. E. Bernstein, A. A. Caudy, S. M. Hammond, G. J. Hannon, *Nature* 409, 363 (2001).
- 25 30. Graham, F.L. and van der Eb, A.J., (1973), *Viol.* 52, 456.
31. McCutchan, J.H. and Pagano, J.S., (1968), *J. Natl. Cancer Inst.* 41, 351.
32. Chu, G. et al., (1987), *Nucl. Acids Res.* 15, 1311.
- 30 33. Fraley, R. et al., (1980), *J. Biol. Chem.* 255, 10431.
34. Capecchi, M.R., (1980), *Cell* 22, 479.
35. Felgner, P.L. et al., (1987), *Proc. Natl. Acad. Sci USA* 84, 7413.

- 25 -

36. Lau N.C., Lim L.P., Weinstein E.G., Bartel D.P., (2001), *Science* 294, 858-862.
37. Lee R.C., Ambros V., (2001), *Science* 294, 862-864.
38. Ambros V., (2001), *Cell* 107, 823-826.
- 5 39. Ambros V., Horvitz H.R., (1984), *Science* 226, 409-416.
40. Wightman B., Ha I., Ruvkun G., (1993), *Cell* 75, 855-862.
41. Rougvie A.E., (2001), *Nat. Rev. Genet.* 2, 690-701.
42. Ketting R.F., Fischer S.E., Bernstein E., Sijen T., Hannon G.J.,  
Plasterk R.H., (2001), *Genes & Dev.* 15, 2654-2659.
- 10 43. Hallam S.J., Jin Y., (1998), *Nature* 395, 78-82.
44. Gauwerky C.E., Huebner K., Isobe M., Nowell P.C., Croce C.M.,  
(1989), *Proc. Natl. Acad. Sci. USA* 86, 8867-8871.
45. P. Chomczynski, N. Sacchi, *Anal Biochem* 162, 156, (1987).
46. Mourelatos Z., Dostie J., Paushkin S., Sharma A., Charroux B., Abel  
15 L., J.R., Mann M., Dreyfuss G., (2002), *Genes & Dev.*, in press.
47. Tam W., (2001), *Gene* 274, 157-167.

**Table 1**

D. melanogaster miRNAs. The sequences given represent the most abundant, and typically longest miRNA sequence identified by cloning; miRNAs frequently vary in length by one or two nucleotides at their 3' termini. From 222 short RNAs sequenced, 69 (31%) corresponded to miRNAs, 103 (46%) to already characterized functional RNAs (rRNA, 7SL RNA, tRNAs), 30 (14%) to transposon RNA fragments, and 20 (10%) sequences with no database entry. The frequency (freq.) for cloning a particular miRNA relative to all identified miRNAs is indicated in percent.

Results of Northern blotting of total RNA isolated from staged populations of D. melanogaster are summarized. E, embryo; L, larval stage; P, pupae; A, adult; S2, Schneider-2 cells. The strength of the signal within each blot is represented from strongest (+ + +) to undetected (-). let-7 stRNA was probed as control. Genbank accession numbers and homologs of miRNAs identified by database searching in other species are provided as supplementary material.

miRNA	sequence (5' to 3')	freq. (%)	E 0-3 h	E 0-6 h	L1+ L2	L3	P	A	S2
miR-1	UGGAAUGUAAAGAAGUAUGGAG (SEQ ID NO:58)	32	+	+	++ +	++ +	++	++ +	-
miR-2a*	UAUCACAGCCAGCUUUGAUGAGC (SEQ ID NO:59)	3							
miR-2b*	UAUCACAGCCAGCUUUGAGGAGC (SEQ ID NO:60)	3	++	++	++	++ +	++	+	++ +
miR-3	UCACUGGGCAAAGUGUGUCUCA#	9	+++	+++	-	-	-	-	-
miR-4	AUAAAGCUAGACAACCAUUGA (SEQ ID NO:62)	6	+++	+++	-	-	-	-	-
miR-5	AAAGGAACGAUCGUUGUGAUAUG (SEQ ID NO:63)	1	+++	+++	+/-	+/-	-	-	-
miR-6	UAUCACAGUGGCUGUUCUUUUU (SEQ ID NO:64)	13	+++	+++	+/-	+/-	-	-	-
miR-7	UGGAAGACUAGUGAUUUUGUUGU (SEQ ID NO:65)	4	+++	++	+/-	+/-	+/-	+/-	+/-
miR-8	UAAUACUGUCAGGUAAAGAUGUC (SEQ ID NO:66)	3	+/-	+/-	++ +	++ +	+	++ +	-

- 27 -

miR-9	UCUUUGGUUAUCUAGCUGUAUGA (SEQ ID NO:67)	7	+++	++	++	++	++	+/	-
miR-10	ACCCUGUAGAUCGAAUUUGU (SEQ ID NO:68)	1	+	+	++	++	+/	+	-
miR-11	CAUCACAGUCUGAGUUCUUGC (SEQ ID NO:69)	7	+++	+++	++	++	++	+	-
miR-12	UGAGUAUUACAUCAGGUACUGGU (SEQ ID NO:70)	7	+	+	++	++	+	++	+/
5 miR-13a*	UAUCACAGCCAUUUUGACGAGU (SEQ ID NO:71)	1	+++	+++	++	++	+	++	++
miR-13b*	UAUCACAGCCAUUUUGAUGAGU (SEQ ID NO:72)	0							
miR-14	UCAGUCUUUUUCUCUCCUA (SEQ ID NO:73)	1	-	-	-	-	-	-	-
let-7	UGAGGUAGUAGGUUGUAUAGUU (SEQ ID NO:74)	0	-	-	-	-	++	++	-

10 # = (SEQ ID NO:61)

\*Similar miRNA sequences are difficult to distinguish by Northern blotting because of potential cross-hybridization of probes.

Table 2

Human miRNAs. From 220 short RNAs sequenced, 100 (45%) corresponded to miRNAs, 53 (24%) to already characterized functional RNAs (rRNA, snRNAs, tRNAs), and 67 (30%) sequences with no database entry. Results of Northern blotting of total RNA isolated from different vertebrate species and S2 cells are indicated. For legend, see Table 1.

	miRNA	sequence (5' to 3')	freq. (%)	HeLa cells	mouse kidney	adult fish	frog ovary	S2
	let-7a*	UGAGGUAGUAGGUUGUAUAGUU#	10	+++	+++	+++	-	-
10	let-7b*	UGAGGUAGUAGGUUGUGUGGUU (SEQ ID NO:76)	13					
	let-7c*	UGAGGUAGUAGGUUGUAUGGUU (SEQ ID NO:77)	3					
	let-7d*	AGAGGUAGUAGGUUGCAUAGU (SEQ ID NO:78)	2	+++	+++	+++	-	-
	let-7e*	UGAGGUAGGAGGUUGUAUAGU (SEQ ID NO:79)	2	+++	+++	+++	-	-
	let-7f*	UGAGGUAGUAGAUUGUAUAGUU (SEQ ID NO:80)	1					
15	miR-15	UAGCAGCACAAUUGGUUGUG (SEQ ID NO:81)	3	+++	++	+	+/-	-
	miR-16	UAGCAGCACGUAUUGGCG (SEQ ID NO:82)	10	+++	+	+/-	+/-	-
	miR-17	ACUGCAGUGAAGGCACUUGU (SEQ ID NO:83)	1	+++	-	-	-	-
	miR-18	UAAGGUGCAUCUAGUGCAGUA (SEQ ID NO:84)	2	+++	-	-	-	-
	miR-19a*	UGUGCAAUUCUUGCAAACUGA (SEQ ID NO:85)	1	+++	-	+/-	-	-
20	miR-19b*	UGUGCAAUCCAUGCAAACUGA (SEQ ID NO:86)	3					
	miR-20	UAAAGUGCUUAGUGCAGGUA (SEQ ID NO:87)	4	+++	-	+	-	-
	miR-21	UAGCUUAUCAGACUGAUGUUGA (SEQ ID NO:88)	10	+++	+	++	-	-
	miR-22	AAGCUGCCAGUUGAAGAACUGU (SEQ ID NO:89)	10	+++	+++	+	+/-	-
	miR-23	AUCACAUUGCCAGGGAUUUCC (SEQ ID NO:90)	2	+++	+++	+++	+	-

- 29 -

	miR-24	UGGCUCAGUUCAGCAGGAACAG (SEQ ID NO:91)	4	++	+++	++	-	-
	miR-25	CAUUGCACUUGUCUCGGUCUGA (SEQ ID NO:92)	3	+++	+	++	-	-
	miR-26a*	UUCAAGUAAUCCAGGAUAGGCU (SEQ ID NO:93)	2	+	++	+++	-	-
	miR-26b*	UUCAAGUAAUUCAGGAUAGGUU (SEQ ID NO:94)	1					-
5	miR-27	UUCACAGUGGCUAAGUCCGCU (SEQ ID NO:95)	2	+++	+++	++	-	-
	miR-28	AAGGAGCUCACAGUCUAUUGAG (SEQ ID NO:96)	2	+++	+++	-	-	-
	miR-29	CUAGCACCAUCUGAAAUCGGUU (SEQ ID NO:97)	2	+	+++	+/-	-	-
	miR-30	CUUUCAGUCGGAUGUUUGCAGC (SEQ ID NO:98)	2	+++	+++	+++	-	-
	miR-31	GGCAAGAUGCUGGCAUAGCUG (SEQ ID NO:99)	2	+++	-	-	-	-
10	miR-32	UAUUGCACAUAUAAGUUGC (SEQ ID NO:100)	1	-	-	-	-	-
	miR-33	GUGCAUUGUAGUUGCAUUG (SEQ ID NO:101)	1	-	-	-	-	-
	miR-1	UGGAUGUAAAGAAGUAUGGAG (SEQ ID NO:102)	0	-	-	+	-	-
	miR-7	UGGAAGACUAGUGAUUUUGUUGU (SEQ ID NO:103)	0	+	-	+/-	-	+/-
	miR-9	UCUUUGGUUAUCUAGCUGUAUGA (SEQ ID NO:104)	0	-	-	-	-	-
15	miR-10	ACCCUGUAGAUCCGAAUUUGU (SEQ ID NO:105)	0	-	+	-	-	-

# = (SEQ ID NO:75)

\*Similar miRNA sequences are difficult to distinguish by Northern blotting because of potential cross-hybridization of probes.

- 30 -

Table 3

Mouse miRNAs. The sequences indicated represent the longest miRNA sequences identified by cloning. The 3'-terminus of miRNAs is often truncated by one or two nucleotides. miRNAs that are more than 85% identical in sequence (i.e. share 18 out of 21 nucleotides) or contain 1- or 2-nucleotide internal deletions are referred to by the same gene number followed by a lowercase letter. Minor sequence variations between related miRNAs are generally found near the ends of the miRNA sequence and are thought to not compromise target RNA recognition. Minor sequence variations may also represent A to G and C to U changes, which are accommodated as G-U wobble base pairs during target recognition. miRNAs with the suffix -s or -as indicate RNAs derived from either the 5'-half or the 3'-half of a miRNA precursor. Mouse brains were dissected into midbrain, mb, cortex, cx, cerebellum, cb. The tissues analyzed were heart, ht; liver, lv; small intestine, si; colon, co; cortex, ct; cerebellum, cb; midbrain, mb.

	miRNA	sequence (5' to 3')	Number of clones							
			ht	lv	sp	si	co	cx	cb	mb
20	let-7a	UGAGGUAGUAGGUUGUAUAGUU (SEQ ID NO:106)	3				1	1		7
	let-7b	UGAGGUAGUAGGUUGUGUGGUU (SEQ ID NO:107)	1	1					2	5
	let-7c	UGAGGUAGUAGGUUGUAUGGUU (SEQ ID NO:108)	2					2	5	19
	let-7d	AGAGGUAGUAGGUUGCAUAGU (SEQ ID NO:109)	2				2	2		2
25	let-7e	UGAGGUAGGAGGUUGUAUAGU (SEQ ID NO:110)			1					2
	let-7f	UGAGGUAGUAGAUUGUAUAGUU (SEQ ID NO:111)			2				3	3
	let-7g	UGAGGUAGUAGUUUGUACAGUA (SEQ ID NO:112)						1	1	2
	let-7h	UGAGGUAGUAGUGUGUACAGUU (SEQ ID NO:113)						1	1	

- 31 -

	let-7i	UGAGGUAGUAGUUUGUGCU (SEQ ID NO:114)				1	1		
	miR-1b	UGGAAUGUAAAGAAGUAUGUAA (SEQ ID NO:115)	4	2					1
	miR-1c	UGGAAUGUAAAGAAGUAUGUAC (SEQ ID NO:116)	7						
	miR-1d	UGGAAUGUAAAGAAGUAUGUAUU (SEQ ID NO:117)	16						1
5	miR-9	UCUUUGGUUAUCUAGCUGUAUGA (SEQ ID NO:118)				3	4		4
	miR-15a	UAGCAGCACAUAAUGGUUUGUG (SEQ ID NO:119)	1						2
	miR-15b	UAGCAGCACAUCAUGGUUUACA (SEQ ID NO:120)	1						
	miR-16	UAGCAGCACGUAAAUUUGGCG (SEQ ID NO:121)	1		1	2	1	2	3
	miR-18	UAAGGUGCAUCUAGUGCAGUA (SEQ ID NO:122)		1					
10	miR-19b	UGUGCAAUCCAUAGCAAACUGA (SEQ ID NO:123)		1					
	miR-20	UAAAGUGCUUAUAGUGCAGGUAG (SEQ ID NO:124)				1			
	miR-21	UAGCUUAUCAGACUGAUGUUGA (SEQ ID NO:125)	1		1	2	1		
	miR-22	AAGCUGCCAGUUGAAGAACUGU (SEQ ID NO:126)	2	1		1		1	2
	miR-23a	AUCACAUUGCCAGGGAUUUCC (SEQ ID NO:127)	1						
15	miR-23b	AUCACAUUGCCAGGGAUUACCAC (SEQ ID NO:128)					1		
	miR-24	UGGCUCAGUUCAGCAGGAACAG (SEQ ID NO:129)	1			1	1		1
	miR-26a	UUCAAGUAAUCCAGGAUAGGCU (SEQ ID NO:130)						3	2
	miR-26b	UUCAAGUAAUCCAGGAUAGGUU (SEQ ID NO:131)		2			4	1	
	miR-27a	UUCACAGUGGCUAAGUUCGCU (SEQ ID NO:132)	1		2		1	1	2
20	miR-27b	UUCACAGUGGCUAAGUUCUG (SEQ ID NO:133)							1
	miR-29a	CUAGCACCAUCUGAAAUCGGUU (SEQ ID NO:134)	1			1		1	
	miR-29b/miR-102	UAGCACCAUUUGAAAUCAGUGUU (SEQ ID NO:135)	1			1	5		3
	miR-29c/	UAGCACCAUUUGAAAUCGGUUA (SEQ ID NO:136)	1				3		1



- 32 -

	miR-30a-s/miR-97	UGUAAACAUCCUCGACUGGAAGC (SEQ ID NO:137)	1	1	1
	miR-30a-as <sup>a</sup>	CUUUCAGUCGGAUGUUUCAGC (SEQ ID NO:138)		1	
	miR-30b	UGUAAACAUCCUACACUCAGC (SEQ ID NO:139)	1	2	
	miR-30c	UGUAAACAUCCUACACUCUCAGC (SEQ ID NO:140)	2	1	1
5	miR-30d	UGUAAACAUCCCCGACUGGAAG (SEQ ID NO:141)	1		
	miR-99a/miR-99	ACCCGUAGAUCCGAUUUGU (SEQ ID NO:142)		1	
	miR-99b	CACCCGUAGAACCGACCUUGCG (SEQ ID NO:143)		1	
	miR-101	UACAGUACUGUGAUAAACUGA (SEQ ID NO:144)		2	1 1
	miR-122a	UGGAGUGUGACAAUGGUGUUUGU (SEQ ID NO:145)	3		
10	miR-122b	UGGAGUGUGACAAUGGUGUUUGA (SEQ ID NO:146)	11		
	miR-122a,b	UGGAGUGUGACAAUGGUGUUUG (SEQ ID NO:147)	23		
	miR-123	CAUUAUUACUUUUGGUACGCG (SEQ ID NO:148)	1 2		
	miR-124a <sup>b</sup>	UUAAGGCACGCGG-UGAAUGCCA (SEQ ID NO:149)	1	37	41 24
	miR-124b	UUAAGGCACGCGGUGAAUGC (SEQ ID NO:150)		1	3
15	miR-125a	UCCUGAGACCCUUUAACCUUG (SEQ ID NO:151)		1	1
	miR-125b	UCCUGAGACCCU--AACUUGUGA (SEQ ID NO:152)		1	
	miR-126	UCGUACCGUGAGUAAUAAUGC (SEQ ID NO:153)	4		1
	miR-127	UCGGAUCCGUCUGAGCUUGGCU (SEQ ID NO:154)			1
	miR-128	UCACAGUGAACCGUCUCUUUU (SEQ ID NO:155)		2	2 2
20	miR-129	CUUUUUUCGGUCUGGGCUUGC (SEQ ID NO:156)			1
	miR-130	CAGUGCAAUGUUAAAAGGGC (SEQ ID NO:157)			1
	miR-131	UAAAGCUAGAUAAACCGAAAGU (SEQ ID NO:158)		1	1 1
	miR-132	UAACAGUCUACAGCCAUGGUCGU (SEQ ID NO:159)			1

- 33 -

	miR-133	UUGGUCCCCUUAACCAGCUGU (SEQ ID NO:160)	4			1		
	miR-134	UGUGACUGGUUGACCAGAGGGA (SEQ ID NO:161)				1		
	miR-135	UAUGGCUUUUUAUCCUAUGUGAA (SEQ ID NO:162)				1		
	miR-136	ACUCCAUUUGUUUGAUGAUGGA (SEQ ID NO:163)				1		
5	miR-137	UAUUGCUUAAGAAUACGCGUAG (SEQ ID NO:164)				1		1
	miR-138	AGCUGGUGUUGUGAAUC (SEQ ID NO:165)				1		
	miR-139	UCUACAGUGCACGUGUCU (SEQ ID NO:166)				1	1	
	miR-140	AGUGGUUUUACCCUAUGGUAG (SEQ ID NO:167)			1			
	miR-141	AACACUGUCUGGUAAGAUGG (SEQ ID NO:168)		1	1		1	
10	miR-142-s	CAUAAAGUAGAAAGCACUAC (SEQ ID NO:169)			1	1		
	miR-142-as <sup>b</sup>	UGUAGUGUUUCCUACUUUAUGG (SEQ ID NO:170)		1	1	6		
	miR-143	UGAGAUGAAGCACUGUAGCUCA (SEQ ID NO:171)	3	7		2		1
	miR-144	UACAGUAUAGAUGAUGUACUAG (SEQ ID NO:172)	2			1		
	miR-145	GUCCAGUUUCCCAGGAAUCCCUU (SEQ ID NO:173)	1					
15	miR-146	UGAGAACUGAAUCCAUGGGUUU (SEQ ID NO:174)	1					
	miR-147	GUGUGUGGAAUGCUUCUGCC (SEQ ID NO:175)		1				
	miR-148	UCAGUGCACUACAGAACUUUGU (SEQ ID NO:176)		1				
	miR-149	UCUGGCUCGUGUCUACUCC (SEQ ID NO:177)	1					
	miR-150	UCUCCCAACCCUUGUACCAGUGU (SEQ ID NO:178)				1		
20	miR-151	CUAGACUGAGGCUCUUUGAGGU (SEQ ID NO:179)				1		
	miR-152	UCAGUGCAUGACAGAACUUGG (SEQ ID NO:180)				1		
	miR-153	UUGCAUAGUCACAAAAGUGA (SEQ ID NO:181)						1
	miR-154	UAGGUUAUCCGUGUUGCCUUCG (SEQ ID NO:182)						1

- 34 -

miR-155

UUAAUGCUAAUUGUGAUAGGGG  
(SEQ ID NO:183)1

---

5 The originally described miR-30 was renamed to miR-30a-as in order to distinguish it from the miRNA derived from the opposite strand of the precursor encoded by the mir-30a gene. miR-30a-s is equivalent to miR-97 [46].

A 1-nt length heterogeneity is found on both 5' and 3' end. The 22-nt miR sequence is shown, but only 21-nt miRNAs were cloned.

10

**Table 4**

Mouse and human miRNAs. The sequences indicated represent the longest miRNA sequences identified by cloning. The 3' terminus of miRNAs is often truncated by one or two nucleotides. miRNAs that are more than 85% identical in sequence (i.e. share 18 out of 21 nucleotides) or contain 1- or 2-nucleotide internal deletions are referred to by the same gene number followed by a lowercase letter. Minor sequence variations between related miRNAs are generally found near the ends of the miRNA sequence and are thought to not compromise target RNA recognition. Minor sequence variations may also represent A to G and C to U changes; which are accommodated as G-U wobble base pairs during target recognition. Mouse brains were dissected into midbrain, mb, cortex, cx, cerebellum, cb. The tissues analyzed were lung, ln; liver, lv; spleen, sp; kidney, kd; skin, sk; testis, ts; ovary, ov; thymus, thy; eye, ey; cortex, ct; cerebellum, cb; midbrain, mb. The human osteosarcoma cells SAOS-2 cells contained an inducible p53 gene (p53-, uninduced p53; p53+, induced p53); the differences in miRNAs identified from induced and uninduced SAOS cells were not statistically significant.

## number of clones

miRNA	Sequence (5' to 3')	mouse tissues										human SAOS-2 cells	
		ln	lv	sp	kd	sk	ts	ov	thy	ey	p53-	p53+	
5													
						1				2			(SEQ ID NO.184)
	miR-C1	1											(SEQ ID NO.185)
10	miR-C2									1			(SEQ ID NO.186)
	miR-C3									1			(SEQ ID NO.187)
	miR-C4						1			1	1		(SEQ ID NO.188)
	miR-C5									2			(SEQ ID NO.189)
	miR-C6									1			(SEQ ID NO.190)
15	miR-C7									1	1		(SEQ ID NO.191)
	miR-C8												(SEQ ID NO.192)
	miR-C9				1								(SEQ ID NO.193)
	miR-C10				1								(SEQ ID NO.194)
	miR-C11				1								(SEQ ID NO.195)
20	miR-C12				2								(SEQ ID NO.196)
	miR-C13			2	1								(SEQ ID NO.197)
	miR-C14	2			1								

miR-C15	UACCACAGGGUAGAACCACGGA	1	(SEQ ID NO.198)
miR-C16	AACUGGCCUACAAAGUCCAG	1	(SEQ ID NO.199)
miR-C17	UGUAAACAGCAACUCCAUGUGA	1	(SEQ ID NO.200)
miR-C18	UAGCAGCACAGAAAUUUGGC	2	(SEQ ID NO.201)
5 miR-C19	UAGGUAGUUUCAUGUUGUUGG	1	(SEQ ID NO.202)
miR-C20	UUCACCAACCUUCUCCACCCAGC	1	(SEQ ID NO.203)
miR-C21	GGUCCAGAGGGGAGAUAGG	1	(SEQ ID NO.204)
miR-C22	CCCAGUGUUCAGACUACCUGUU	1	(SEQ ID NO.205)
miR-C23	UAAUACUGCCUGGUAUGAUGAC	2	(SEQ ID NO.206)
10 miR-C24	UACUCAGUAAGGCAUUGUUCU	1	(SEQ ID NO.207)
miR-C25	AGAGGUUAUAGCGCAUGGGAAGA	1	(SEQ ID NO.208)
miR-C26	UGAAAUUGUUUAGGACCACUAG	1	(SEQ ID NO.209)
miR-C27	UUCUUUUGUCAUCCUAUGCCUG	1	(SEQ ID NO.210)
miR-C28	UCCUUCAUUCCACCGGAGUCUG	1	(SEQ ID NO.211)
15 miR-C29	GUGAAAUUGUUUAGGACCACUAGA	2	(SEQ ID NO.212)
miR-C30	UGGAAUGUAAGGAAGUGUGUGG	2	(SEQ ID NO.213)
miR-C31	UACAGUAGUCUGCACAUUGGUU	1	(SEQ ID NO.214)
miR-C32	CCCUGUAGAACCGAAUUUGUGU	1	(SEQ ID NO.215)
miR-C33	AACCCGUAGAUCCGAACUUGUGAA	1	(SEQ ID NO.216)
20 miR-C34	GCUUCUCCUGGCUCUCCUCCUC	1	(SEQ ID NO.217)

Table 5

*D. melanogaster* miRNA sequences and genomic location. The sequences given represent the most abundant, and typically longest miRNA sequences identified by cloning. It was frequently observed that miRNAs vary in length by one or two nucleotides at their 3'-terminus. From 222 short RNAs sequenced; 69 (31%) corresponded to miRNAs, 103 (46%) to already characterized functional RNAs (rRNA, 7SL RNA, tRNAs), 30 (14%) to transposon RNA fragments, and 20 (10%) sequences with no database entry. RNA sequences with a 5'-guanosine are likely to be underrepresented due to the cloning procedure (8). miRNA homologs found in other species are indicated. Chromosomal location (chr.) and GenBank accession numbers (acc. nb.) are indicated. No ESTs matching miR-1 to miR-14 were detectable by database searching.

miRNA	sequence (5' to 3')	chr., acc. nb.	remarks
miR-1	UGGAAUGTAAAGAAGUAUGGAG (SEQ ID NO: 58)	2L, AE003667	homologs: <i>C. briggsae</i> , G20U, AC87074; <i>C. elegans</i> G20U, U97405; mouse, G20U, G22U, AC020867; human, chr. 20, G20U, G22U, AL449263; ESTs: zebrafish, G20U, G22U, BF157-601; cow, G20U, G22U, BE722-224; human, G20U, G22U, AI220268
miR-2a	UAUCACAGCCAGCUUUGAUGAGC (SEQ ID NO: 59)	2L, AE003663	2 precursor variants clustered with a copy of <i>mir-2b</i>
miR-2b	UAUCACAGCCAGCUUUGAGGAGC (SEQ ID NO: 60)	2L, AE003620 2L, AE003663	2 precursor variants
miR-3	UCACUGGGCAAAGUGUGUCUCA (SEQ ID NO: 61)	2R, AE003795	in cluster <i>mir-3</i> to <i>mir-6</i>
miR-4	AUAAAGCUAGACAACCAUUGA (SEQ ID NO: 62)	2R, AE003795	in cluster <i>mir-3</i> to <i>mir-6</i>

- 39 -

	miR-5	AAAGGAACGAUCGUUGUGAUAUG (SEQ ID NO:63)	2R, AE003795	in cluster <i>mir-3</i> to <i>mir-6</i>
	miR-6	UAUCACAGUGGCUGUUCUUUUU (SEQ ID NO:64)	2R, AE003795	in cluster <i>mir-3</i> to <i>mir-6</i> with 3 variants
5	miR-7	UGGAAGACUAGUGAUUUUGUUGU (SEQ ID NO:65)	2R, AE003791	homologs: human, chr. 19 AC006537, EST BF373391; mouse chr. 17 AC026385, EST AA881786
	miR-8	UAAUACTUGUCAGGUAAGAUGUC (SEQ ID NO:66)	2R, AE003805	
10	miR-9	UCUUUGGUUAUCUAGCUGUAUGA (SEQ ID NO:67)	3L, AE003516	homologs: mouse, chr. 19, AF155142; human, chr. 5, AC026701, chr. 15, AC005316
	miR-10	ACCCUGUAGAUCCGAAUUUGU (SEQ ID NO:68)	AE001574	homologs: mouse, chr 11, AC011194; human, chr. 17, AF287967
	miR-11	CAUCACAGUCUGAGUUCUUGC (SEQ ID NO:69)	3R, AE003735	intronic location
15	miR-12	UGAGUAUUACAUCAGGUACUGGU (SEQ ID NO:70)	X, AE003499	intronic location
	miR-13a	UAUCACAGCCAUUUUGACGAGU (SEQ ID NO:71)	3R, AE003708 X, AE003446	<i>mir-13a</i> clustered with <i>mir-13b</i> on chr. 3R
20	miR-13b	UAUCACAGCCAUUUUGAUGAGU (SEQ ID NO:72)	3R, AE003708	<i>mir-13a</i> clustered with <i>mir-13b</i> on chr. 3R
	miR-14	UCAGUCUUUUUCUCUCUCCUA (SEQ ID NO:73)	2R, AE003833	no signal by Northern analysis

---



- 40 -

Table 6

Human miRNA sequences and genomic location. From 220 short RNAs sequenced, 100 (45%) corresponded to miRNAs, 53 (24%) to already characterized functional RNAs (rRNA, snRNAs, tRNAs), and 67 (30%) sequences with no database entry. For legend, see Table 1.

miRNA	sequence (5' to 3')	chr. or EST, acc. nb.	remarks*
10 let-7a	UGAGGUAGUAGGUUGUAUAGUU (SEQ ID NO:75)	9, AC007924, 11, AP001359, 17, AC087784, 22, AL049853	sequences of chr 9 and 17 identical and clustered with <i>let-7f</i> , homologs: <i>C. elegans</i> , AF274345; <i>C. briggsae</i> , AF210771, <i>D. melanogaster</i> , AE003659
let-7b	UGAGGUAGUAGGUUGUGUGGUU (SEQ ID NO:76)	22, AL049853†, ESTs, AI382133, AW028822	homologs: mouse, EST AI481799; rat, EST, BE120662
let-7c	UGAGGUAGUAGGUUGUAUGGUU (SEQ ID NO:77)	21, AP001667	Homologs: mouse, EST, AA575575
15 let-7d	AGAGGUAGUAGGUUGCAUAGU (SEQ ID NO:78)	17, AC087784, 9, AC007924	identical precursor sequences
let-7e	UGAGGUAGGAGGUUGUAUAGU (SEQ ID NO:79)	19, AC018755	
20 let-7f	UGAGGUAGUAGAUGUAUAGUU (SEQ ID NO:80)	9, AC007924, 17, AC087784, X, AL592046	sequences of chr 9 and 17 identical and clustered with <i>let-7a</i>
miR-15	UAGCAGCACAAUUGGUUUGUG (SEQ ID NO:81)	13, AC069475	in cluster with <i>mir-16</i> homolog
miR-16	UAGCAGCACGUAAUAUUGGCG (SEQ ID NO:82)	13, AC069475	in cluster with <i>mir-15</i> homolog

- 41 -

	miR-17	ACUGCAGUGAAGGCACUUGU (SEQ ID NO:83)	13, AL138714	in cluster with <i>mir-17</i> to <i>mir-20</i>
	miR-18	UAAGGUGCAUCUAGUGCAGATA (SEQ ID NO:84)	13, AL138714	in cluster with <i>mir-17</i> to <i>mir-20</i>
5	miR-19a	UGUGCAAAUCUAUGCAAAACUG A (SEQ ID NO:85)	13, AL138714	in cluster with <i>mir-17</i> to <i>mir-20</i>
	miR-19b	UGUGCAAAUCCAUGCAAAACUG A (SEQ ID NO:86)	13, AL138714, X, AC002407	in cluster with <i>mir-17</i> to <i>mir-20</i>
	miR-20	UAAAGUGCUUAUAGUGCAGGUA (SEQ ID NO:87)	13, AL138714	in cluster with <i>mir-17</i> to <i>mir-20</i>
10	miR-21	UAGCUUAUCAGACUGAUGUUGA (SEQ ID NO:88)	17, AC004686, EST, BF326048	homologs: mouse, EST, AA209594
	miR-22	AAGCUGCCAGUUGAAGAACUGU (SEQ ID NO:89)	ESTs, AW961681†, AA456477, AI752503, BF030303, HS1242049	human ESTs highly similar; homologs: mouse, ESTs, e.g. AA823029; rat, ESTs, e.g. BF543690
15	miR-23	AUCACAUTGCCAGGGAUUTCC (SEQ ID NO:90)	19, AC020916	homologs: mouse, EST, AW124037; rat, EST, BF402515
	miR-24	UGGCUCAGUUCAGCAGGAACAG (SEQ ID NO:91)	9, AF043896, 19, AC020916	homologs: mouse, ESTs, AA111466, AI286629; pig, EST, BE030976
20	miR-25	CAUUGCACUUGUCUGGUCUGA (SEQ ID NO:92)	7, AC073842, EST, BE077684	human chr 7 and EST identical; highly similar precursors in mouse ESTs (e.g. AI595464); fish precursor different STS: G46757
	miR-26a	UUCAAGTAAUCCAGGAUAGGCU (SEQ ID NO:93)	3, AP000497	

- 42 -

	miR-26b	UUCAAGUAAUUCAGGAUAGGUU (SEQ ID NO:94)	2, AC021016	
	miR-27	UUCACAGUGGCUAAGUCCGCTU (SEQ ID NO:95)	19, AC20916	U22C mutation in human genomic sequence
5	miR-28	AAGGAGCUCACAGUCUAUUGAG (SEQ ID NO:96)	3, AC063932	
	miR-29	CUAGCACCAUCUGAAAUCGGUU (SEQ ID NO:97)	7, AF017104	
	miR-30	CUUUCAGUCGGAUGUUUGCAGC (SEQ ID NO:98)	6, AL035467	
10	miR-31	GGCAAGAUGCUGGCAUAGCUG (SEQ ID NO:99)	9, AL353732	
	miR-32	UAUUGCACAUUACTAAGUUGC (SEQ ID NO:100)	9, AL354797	not detected by Northern blotting
15	miR-33	GUGCAUUGUAGUUGCAUUG (SEQ ID NO:101)	22, Z99716	not detected by Northern blotting

\*If several ESTs were retrieved for one organism in the database, only those with different precursor sequences are listed.

20 †precursor structure shown in Fig. 4.

### Claims

1. Isolated nucleic acid molecule comprising
  - (a) a nucleotide sequence as shown in Table 1, Table 2, Table 3 or Table 4 or a precursor thereof as shown in Figure 3, Figure 4 or Figure 7.
  - (b) a nucleotide sequence which is the complement of (a),
  - (c) a nucleotide sequence which has an identity of at least 80% to a sequence of (a) or (b) and/or
  - (d) a nucleotide sequence which hybridizes under stringent conditions to a sequence of (a), (b) and/or (c).
2. The nucleic acid molecule of claim 1, wherein the identity of sequence (c) is at least 90%.
3. The nucleic acid molecule of claim 1, wherein the identity of sequence (c) is at least 95%.
4. The nucleic acid molecule of any one of claims 1-3, which is selected from miR 1-14 as shown in Table 1 or miR 15-33 as shown in Table 2 or miR 1-155 as shown in Table 3 or miR-C1-34 as shown in Table 4 or a complement thereof.
5. The nucleic acid molecule of any one of claims 1-3, which is selected from mir 1-14 as shown in Figure 3 or let 7a-7f or mir 15-33, as shown in Figure 4 or let 7a-i or mir 1-155 or mir-c1-34, as shown in Figure 7 or a complement thereof.

- 44 -

6. The nucleic acid molecule of any one of claims 1-4 which is a miRNA molecule or an analog thereof having a length of from 18-25 nucleotides.
7. The nucleic acid molecule of any one of claims 1-3 or 5, which is a  
5 miRNA precursor molecule having a length of 60-80 nucleotides or a DNA molecule coding therefor.
8. The nucleic acid molecule of any one of claims 1-7, which is single-stranded.
- 10 9. The nucleic acid molecule of any one of claims 1-7, which is at least partially double-stranded.
- 15 10. The nucleic acid molecule of any one of claims 1-9, which is selected from RNA, DNA or nucleic acid analog molecules.
11. The nucleic acid molecule of claim 10, which is a molecule containing at least one modified nucleotide analog.
- 20 12. The nucleic molecule of claim 10 which is a recombinant expression vector.
13. A pharmaceutical composition containinig as an active agent at least one nucleic acid molecule of any one of claims 1-12 and optionally a pharma-  
25 ceutically acceptable carrier.
14. The composition of claim 13 for diagnostic applications.
15. The composition of claim 13 for therapeutic applications.
- 30 16. The composition of any one of claims 13-15 as a marker or a modulator for developmental or pathogenic processes.

- 45 -

17. The composition of claim 13 as a marker or modulator of developmental disorders, particularly cancer, such a B-cell chronic leukemia.
18. The composition of any one of claims 13-15 as a marker or modulator of gene expression.  
5
19. The composition of claim 18 as a marker or modulator of the expression of a gene, which is at least partially complementary to said nucleic acid molecule.  
10
20. A method of identifying microRNA molecules or precursor molecules thereof comprising ligating 5'- and 3'-adapter molecules to the ends of a size-fractionated RNA population, reverse transcribing said adapter-containing RNA population and characterizing the reverse transcription products.  
15

Fig. 1 A

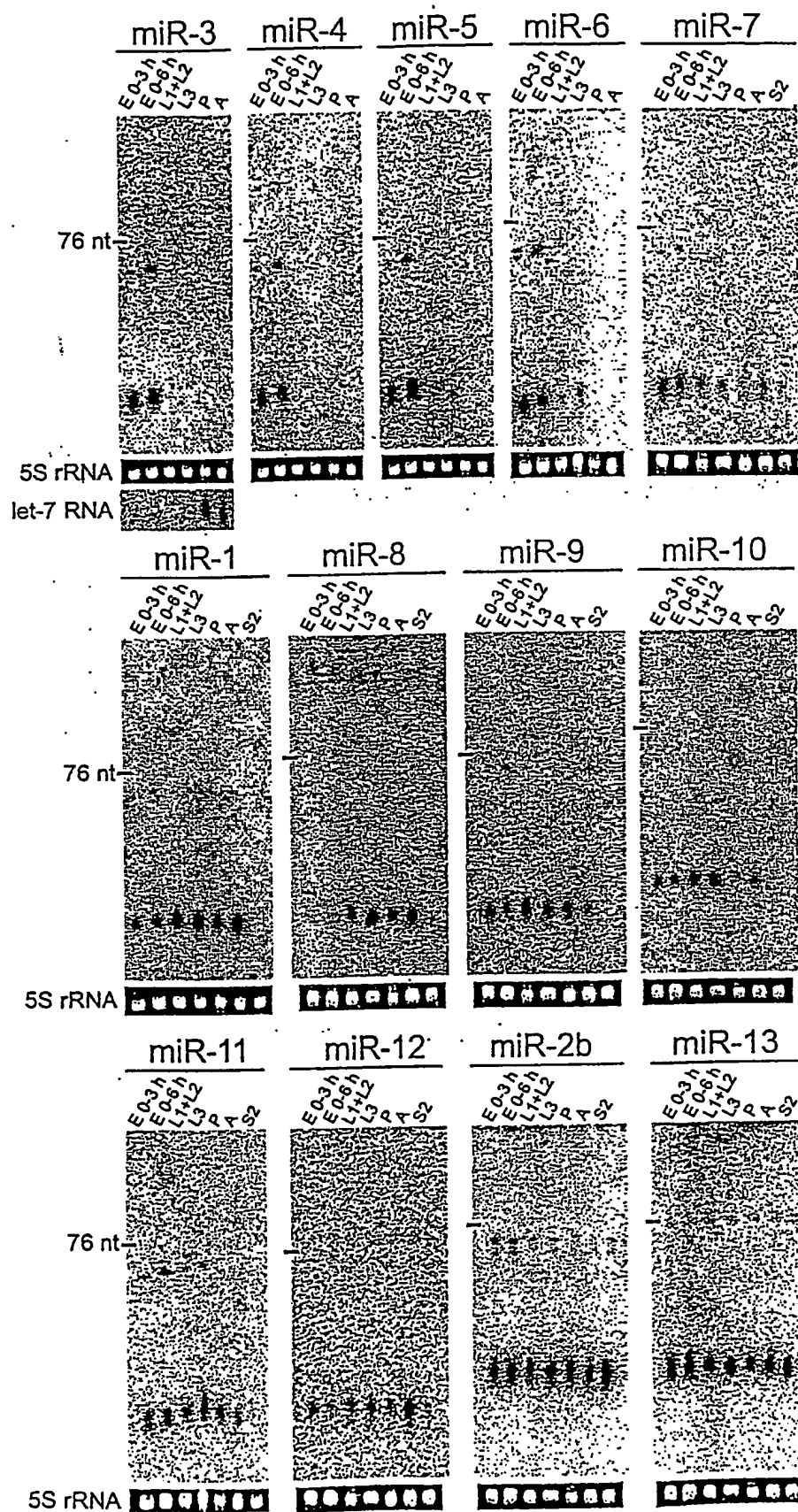


Fig. 1 B

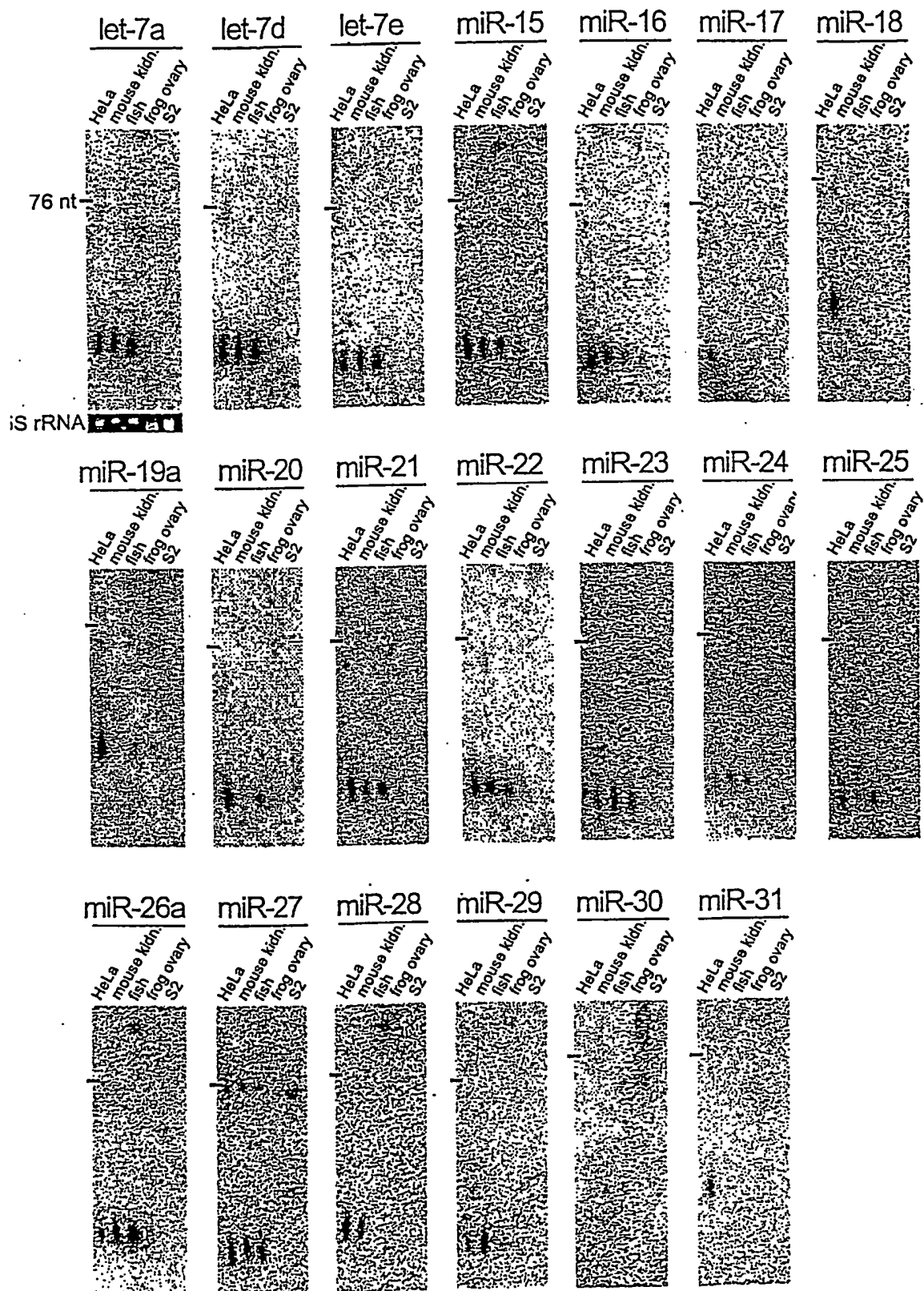




Fig. 2

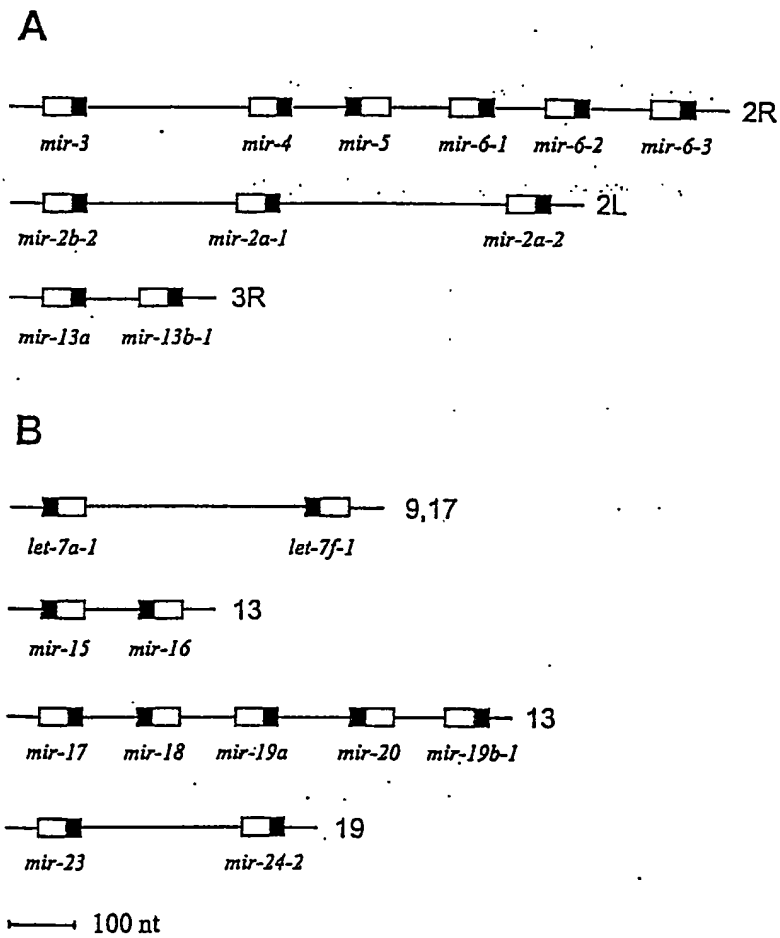


Fig. 3

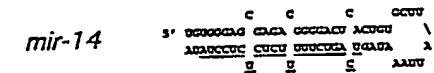
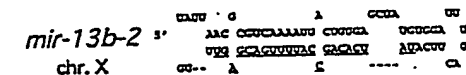
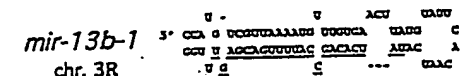
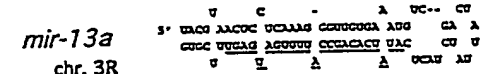
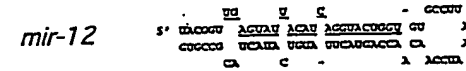
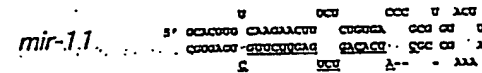
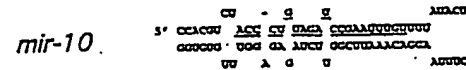
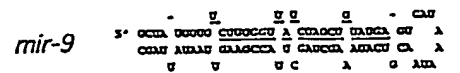
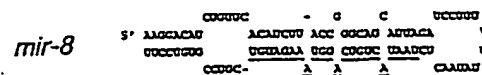
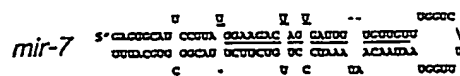
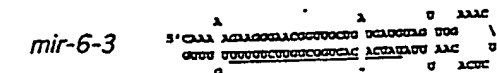
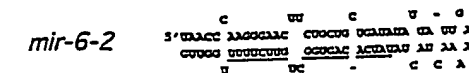
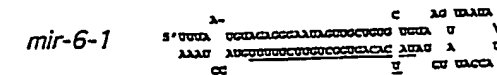
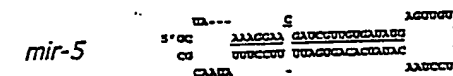
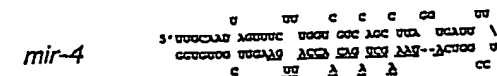
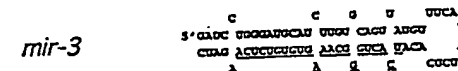
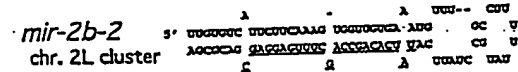
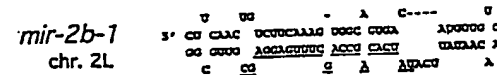
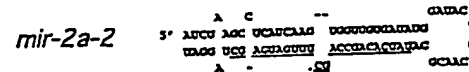
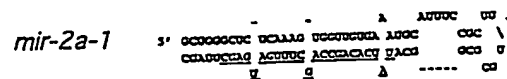
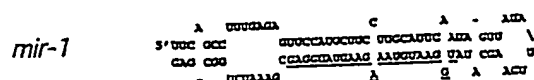


Fig. 4

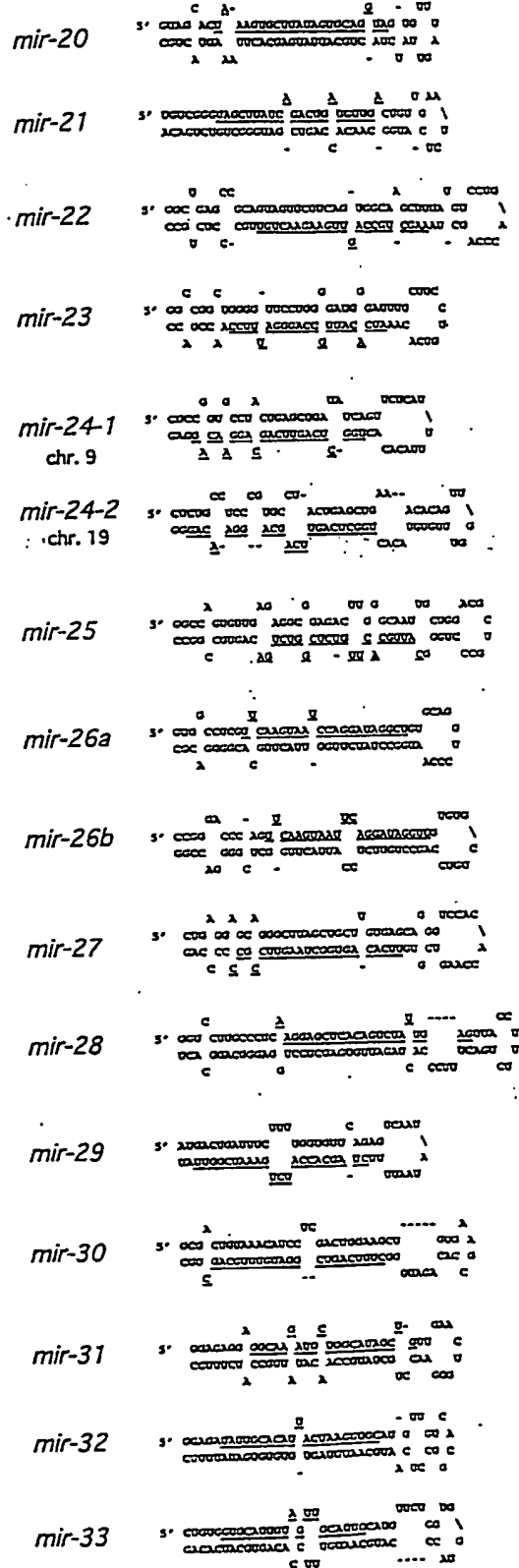
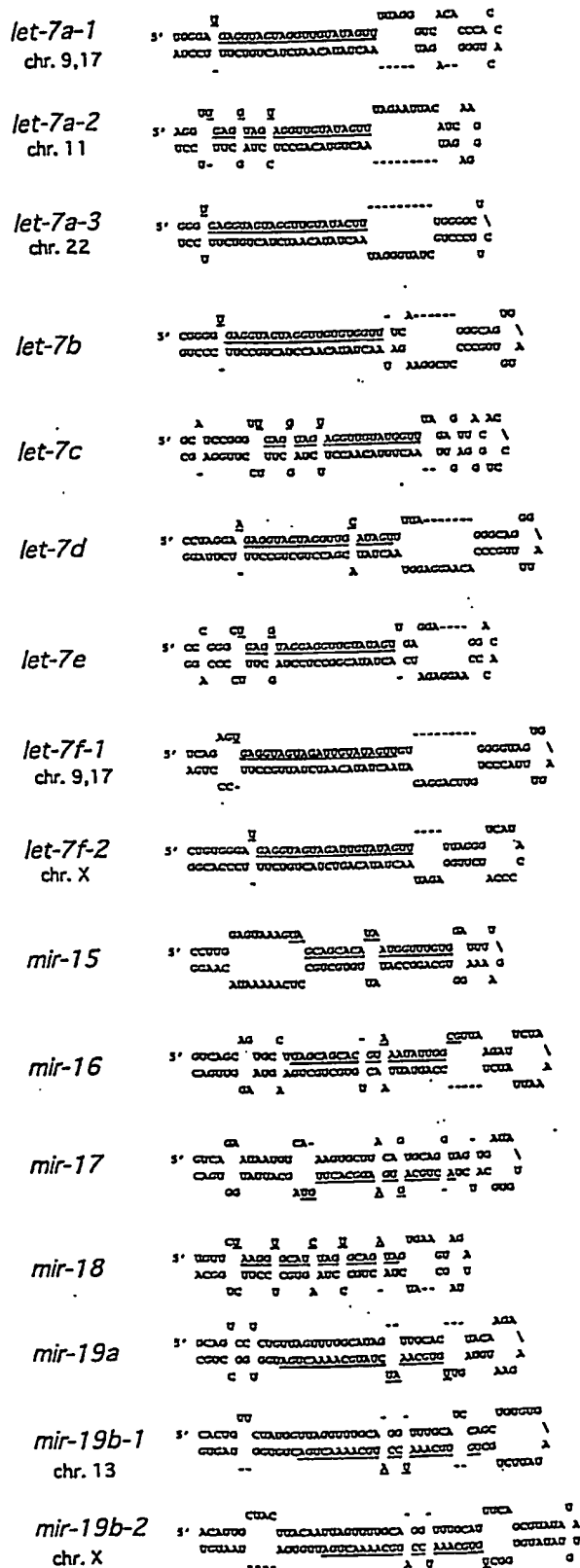


Fig. 5

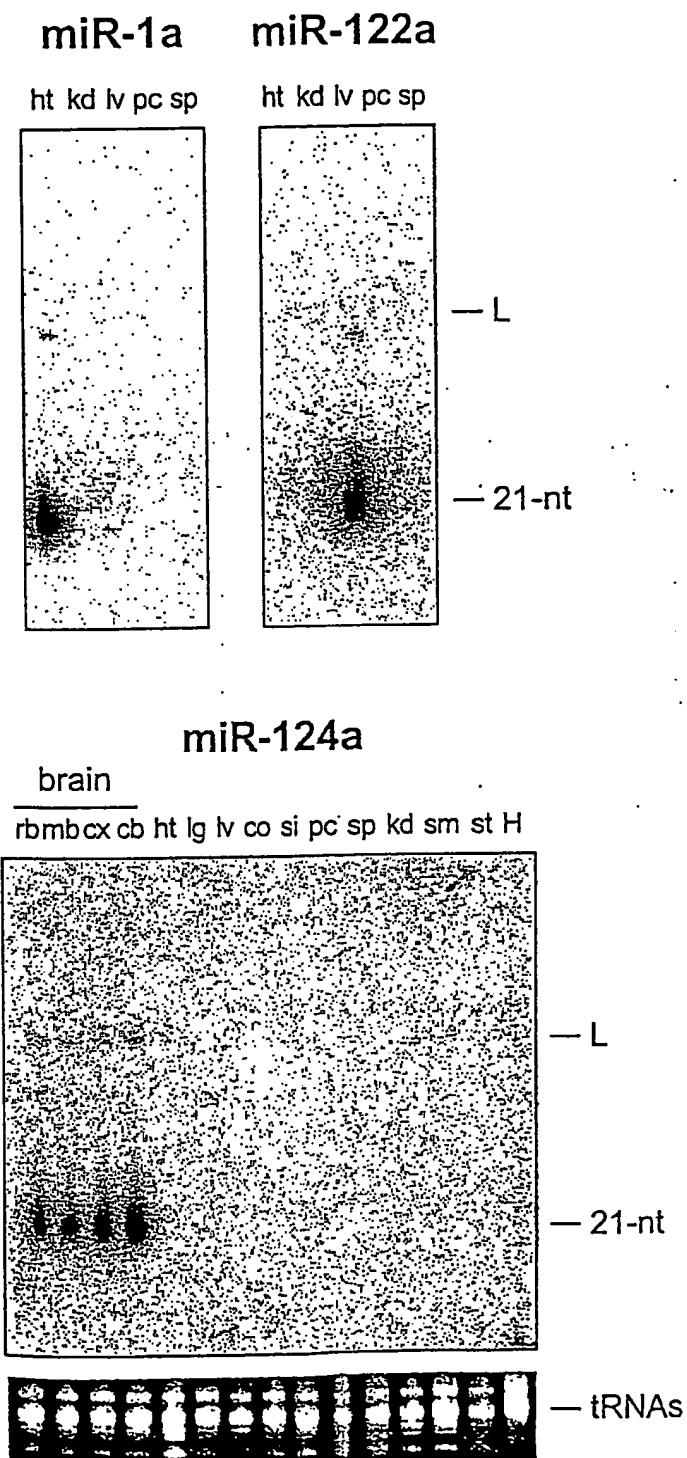


Fig. 5 (cont.)

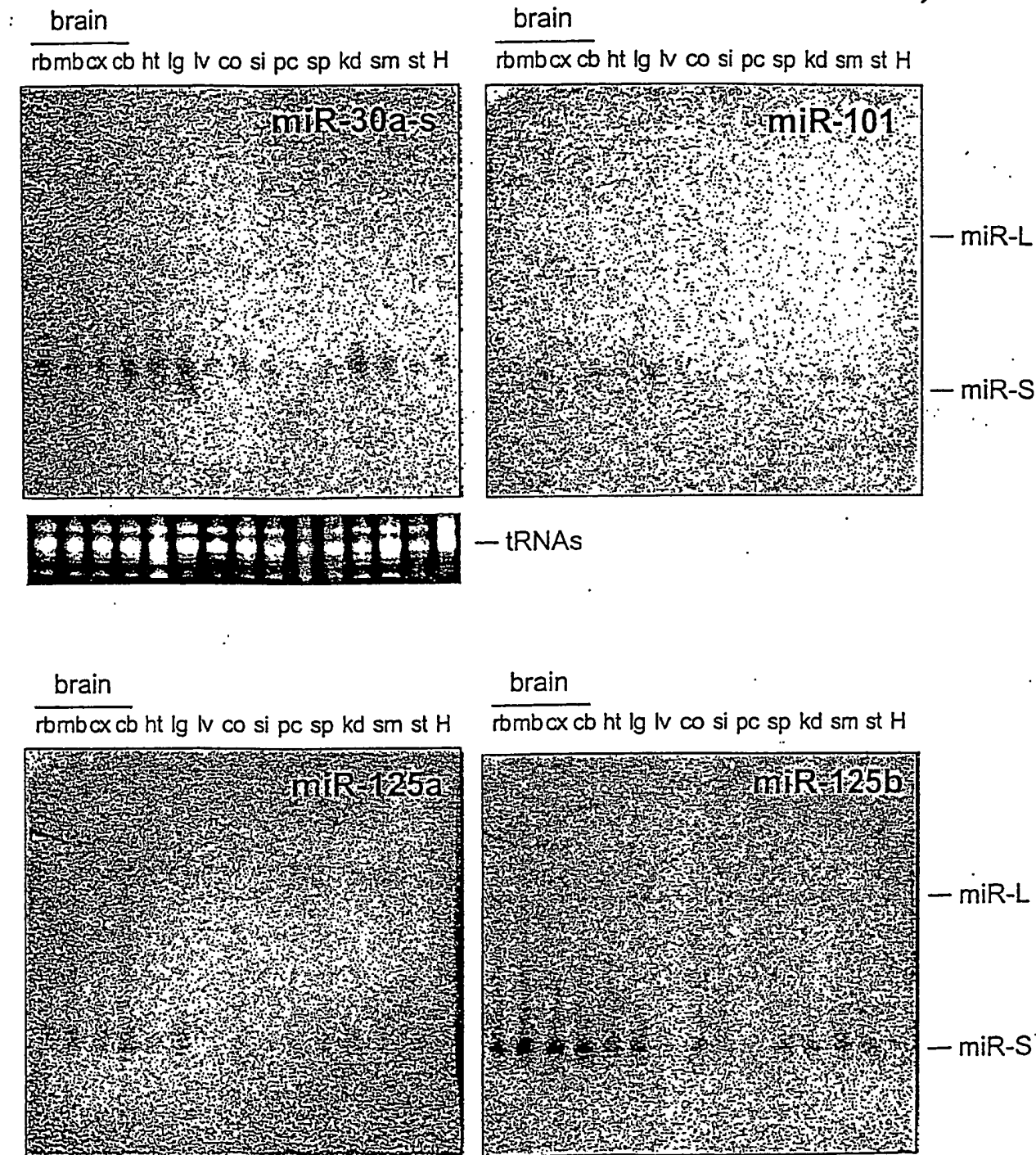


Fig. 5 (cont.)

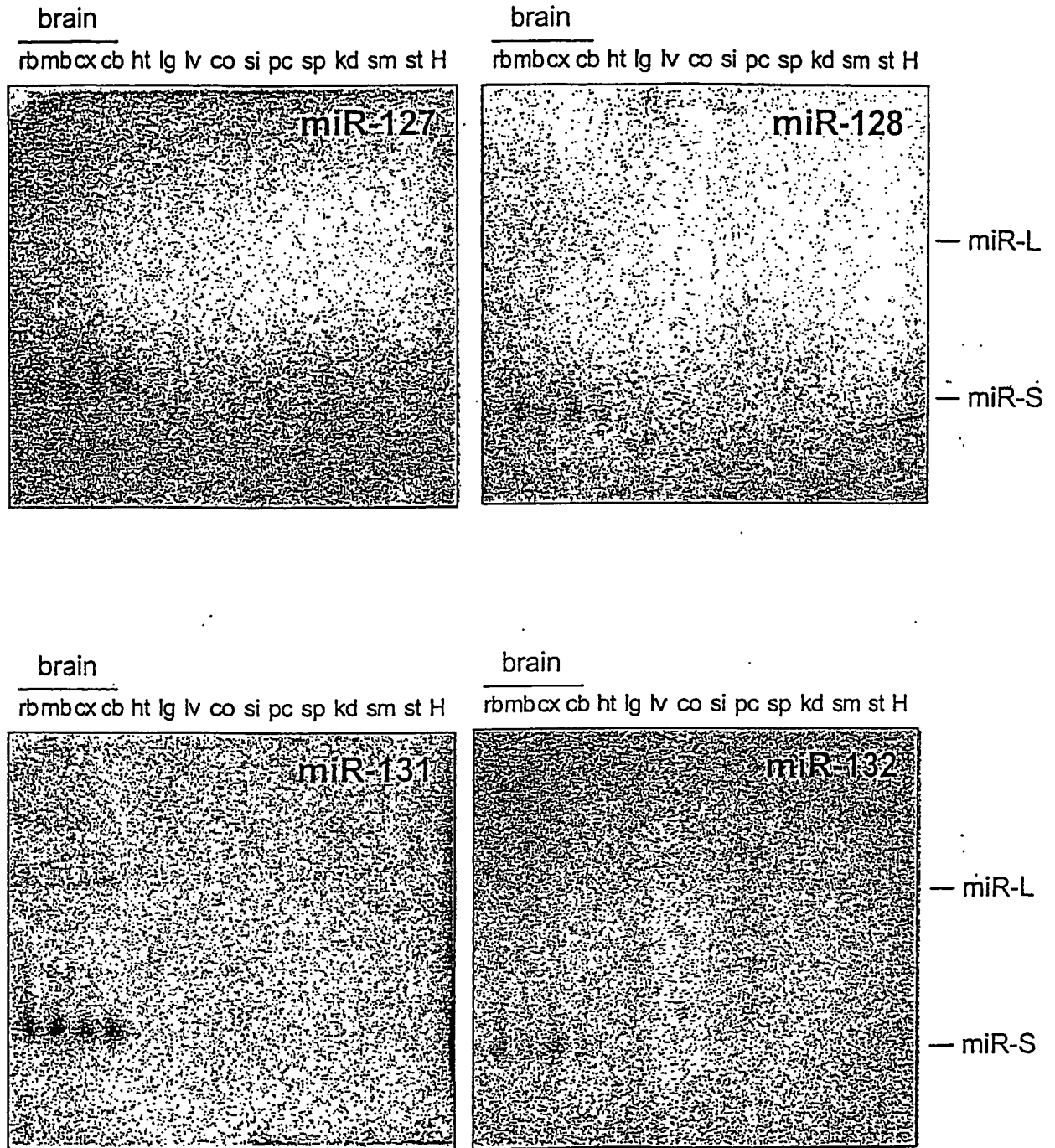


Fig. 5 (cont.)

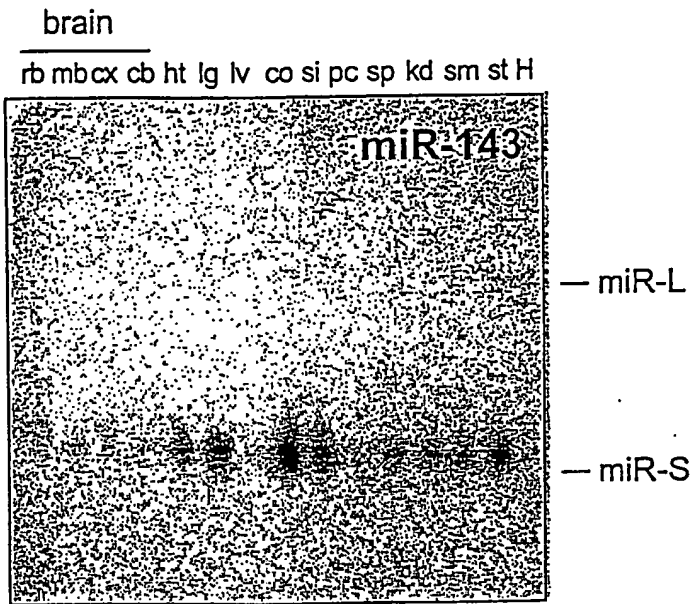


Fig. 6

A

*C. elegans* lin-4*D. melanogaster* miR-125*M. musculus*/H. sapiens miR-125b*M. musculus*/H. sapiens miR-125a

UCCCUAGAGACCUC--AAG-UGUGA

UCCCUAGAGACCCU--AAUUGUGA

UCCCUAGAGACCCU--AAUUGUGA

UCCCUAGAGACCCUUAACUUGUGA

B

0-3 0-6 L1+L2 L3 2 1 3 2





Fig. 7

name	sequence	structure
let-7a-1	UGAGGUAGUAGGUUGUAUAGUU	UG <u>U</u> UGAGG ACA C CAC UGGGA GAGGUAGGUUGUAUAGUU GUC CCCA C GUG AUCCU UUCUGUCAUCAUAUCAA UAG GGU A CA - - - - - A - - C
let-7a-2	UGAGGUAGUAGGUUGUAUAGUU	UU <u>G</u> U UAGAUAUAC AA AGG GAG UAG AGGUUGUAUAGUU AUC G UCC UUC AUC UCCGACAUGUCAA UAG G U- G C - - - - - AG
let-7a-3	UGAGGUAGUAGGUUGUAUAGUU	U GAGGUAGUAGGUUGUAUAGUU U GGG GAGGUAGUAGGUUGUAUAGUU UGGGC \ U UCC UUCUGUCAUCAUAUCAA UAGGUUUC GUCCC G C U
let-7b	UGAGGUAGUAGGUUGUGUGGUU	GG <u>U</u> - - A - - - - - UG CGGG GAGGUAGUAGGUUGUGUGGUU UC GGCAG \ GUCCC UUCGUCAUCCAACAUAUCAA AG CCCGUU A - - - - - U AAGGCU GU
let-7c	UGAGGUAGUAGGUUGUAUGGUU	A UU <u>G</u> U UA G UA AC GC UCCGGG GAG UAG AGGUUGUAUGGUU GA U C \ CG AGGUUC UUC AUC UCCAACAUGUCAA UU A G C - CU G U - - G GG UC
let-7d	AGAGGUAGUAGGUUGCAUAGU	A A GAGGUAGUAGGUUG AUAUU GGCAG \ GG CCUAGGA GAGGUAGUAGGUUG AUAUU CCCGUU A GGAUUCU UUCGUCGUCCAGC UAUCAA UUGAGGAACA UU
let-7e	UGAGGUAGGAGGUUGUAUAGU	C CU <u>G</u> U GGA - - - - - A CC GGG GAG UAGGAGGUUGUAUAGU GA GG C GG CCC UUC AUCCUCCGGCAUAUCA CU CC A A CU G - - AGAGGAA C

Fig. 7 (cont.)

let-7f-1	UGAGGUAGUAGAUUGUAUAGUU	AGU UCAG GAGGUAGUAGAUUGUAUAGUUUGU \ UG AGUC UUCGGUUAUCUAACAUAUCAUA GGGUAG \ UCCCAUU A CC- GAGGACUUG UU
let-7f-2	UGAGGUAGUAGAUUGUAUAGUU	U CUGUGGGA GAGGUAGUAGAUUGUAUAGUU UCAU GGCACCCU UUCUGUCAUCUGACAUAUCA UAGG A UAGA ACCC C
let-7g	UGAGGUAGUAGUUUGUACAGUA	A U CC GGC GAGGUAGU GUUUGUACAGUU UGAGG A- A A GG CCG UUCCGUCA CCGACAUGUCA GUCU UG UACC C A - C ----- GG - C
let-7h	UGAGGUAGUAGUGUGUACAGUU	
let-7i	UGAGGUAGUAGUUUGUGCU	U CUGGC GAGGUAGUAGUUUGUGC GUU U UGUG GAUCG UUCCGUCAUGGACGCG CAA UC GCGCG A U UAGAGGUG - UUAC
mir-1	UGGAAUGUAAGAAGUAUGGAG	A UUUGAGA C A - AUA UUC GCC GUUCCAUGCUUC UUGCAUUC AUA GUU \ GAG CGG CGAGGUAAGAAG AAUGUAAG UAU CGA U - UCUAAG A G A ACU
mir-1b	UGGAAUGUAAGAAGUAUGUAA	A UGGGA ACAUACUUCUUUAUAU GC ---- AC ACUCU UGUUAUGAAGAAUGUA GGUAU AUC C A- CGA GU AL449263.5

Fig. 7 (cont.)

miR-1c	UGGAAUGUAAAGAAGUAUGUAC	
miR-1d	UGGAAUGUAAAGAAGUAUGUAUU	<p>C GC UGAACC</p> <p>GCUUGGGA ACAUACUUCUUUAUUAU CCAUA U</p> <p>CGGACUUU UGUUAUGAAGAAUAUGUA A- CGUAUC</p>
miR-2a-1	UAUCACAGCCAGCUUUGAUGAGC	<p>- - - A AUUUC UU</p> <p>GCUGGGCUC UCAAAAG UGGUUGUGA AUGC CGC \</p> <p>CGAUUCGAG AGUUUC ACCGACACU UACG GCG U</p> <p>U G A ----- CG</p>
miR-2a-2	UAUCACAGCCAGCUUUGAUGAGC	<p>A C -- -- GAUAC</p> <p>AUCU AGC UCAUCAAG UGGUUGUGUAUUG \</p> <p>UAGG UCG AGUAGUUU ACCGACACUAUAC C</p> <p>A - - CG GCAAC</p>
miR-2b-1	UAUCACAGCCAGCUUUGAUGAGC	<p>U UG - - A C---- U</p> <p>CU CAAC UCUCUCAAAG UGGC GUGA AUGUUG C</p> <p>GG GUUG AGGAGUUUC ACCG CACU UUAUAC A</p> <p>C CG G A AUACU A</p>
miR-2b-2	UAUCACAGCCAGCUUUGAUGAGC	<p>A A UUU--- CUU</p> <p>UUGUGUC UUCUUCAAAG UGGUUGUGA AUG GC U</p> <p>AGCGCAG GAGGAGUUUC ACCGACACU UAC CG U</p> <p>C G A UUAUC UAU</p>
miR-3	UCACUGGGCAAAGUGUGUCUCA	<p>C C C G U UUCA</p> <p>GAUC UGGGAUGCAU UUGU CAGU AUGU \</p> <p>CUAG ACUCUGUGUG AACG GUCA UACA A</p> <p>A A G C CUCU</p>



Fig. 7 (cont.)

miR-9	UCUUUGGUUAUCUAGCUGUAUGA	- GCUA UGUUG CUUUGGU UAU G - GAU CGAU AUAUU GAAGCCA GAUCGA AUACU CA A U U UUC A G AUA
miR-10	ACCCUGUAGAUCCGAAUUUGU	CU - G U AUACU CCACGU ACC CU UAGA CCGAAUUUGUUU A GGUGUG UGG GA AUUU GGCUAAAACAGGA G UU A G U AUUUC
miR-11	CAUCACAGUCUGAGUUCUUGC	U UCU CCC U ACU GCACUUG CAAGAACUU CUGUGA GCG GU U CGUGAGU GUUCUUGAG GACACU CGC CG . A C UCU A-- - AAA
miR-12	UGAGUAUUACAUCAGGUACUGGU	UG U C - GCCUU UACGGU AGUAU ACAU AGGUACUGGU GU A GUGCCG UCAUA UGUA UUCAUGACCA CA . A CA C - A ACCUA
miR-13a	UAUCACAGCCAUUUUGAUGAGU	U C - A UC-- CU UACG AACUC UCAAAG GGUUGUGA AUG GA A GUGC UUGAG AGUUUU CCGACACU UAC CU U U U A A UCAU AU
miR-13b-1	UAUCACAGCCAUUUUGACGAGU	UG-- U ACU UAUU CCA UCGUAAAAAUG UUGUGA UAUG C GGU AGCAGUUUUAC GACACU AUAC A UUG C --- URAC
miR-13b-2	UAUCACAGCCAUUUUGACGAGU	UAUU G A GCUA UU AAC CGUCAAAAUG CUGUGA UGUGGA U UUG GCAGUUUUUAC GACACU AUACU G GU-- A C ---- CA

Fig. 7 (cont.)

miR-14	UCAGUCUUUUUCUCUCUCCUA	C C C C GCUU UGUGGGAG GAGA GGGGACU ACUGU \ AUAUCCUC CUCU UUUUCUGA UGAUA A U U U C AAUU
miR-15a	UAGCAGCACAUAAUGGUUUGUG	GAGUAAAGUA UA GA U CCUUG GCAGCACAU AUGGUUUGU UUU \ GGAAC CGUCGUGU UACCGGACGU AAA G AUA AAAACUC UA GG A
miR-15b	UAGCAGCACAUCAUGGUUUACA	U C C C A A ACA CUG AGCAGCA AU AUGGUUU CAU CU \ GAU UCGUCGU UA UACUAAG GUA GA G C U U U C - ACU
miR-16	UAGCAGCACGUAUAUUAUUGGCG	AG C - A CGUUA UCUA GUCAGC UGC UUAGCAGCAC GU AAUAUUGG AGAU \ CAGUUG AUG AGUCGUCGUG CA UUAUGACC UCUA A GA A U A ----- UUA
miR-16	only different precursor	UC CU UA C AG AAU GU CACU AGCAGCACG AAUAUUGG GU UGA A CA GUGA UCGUCGUGU UUAUAACC CA AUU U GU UU CA A A... AUA
miR-17	ACUGCAGUGAAGGCACUUGU	GA CA- A G G - AUA GUCA AUAUUGU AAGUGCUU CA UGCAG UAG UG \ CAGU UAUUACG UUCACGGA GU ACGUC AUC AC U GG AUG A G - U GUG
miR-18	UAAGGUGCAUCUAGUCAGAU	CU U C U A UGAA AG UGUU AAGG GCAU UAG GCAG UAG GU A ACGG UUCG CGUG AUC CGUC AUC CG U UC U A C - UA-- AU



Fig. 7 (cont.)

miR-23b	AUCACAUUGCCAGGGAUUAACCAC	<p>C U -- C GUGACU</p> <p>GG UGC UGG GUUCCUGGCA UG UGAUUU U</p> <p>CC ACG ACC UAGGGACCGU AC ACUAAA G</p> <p>A C AU U - AUUAGA</p>
miR-24-1	UGGCUCAGUUCAGCAGGAACAG	<p>G G A UA UCUCAU</p> <p>CUCC GU CCU CUGAGCUGA UCAGU \</p> <p>GAGG CA GGA GACUUGACU GGUCA U</p> <p>A A C C- CACAUU</p>
miR-24-2	UGGCUCAGUUCAGCAGGAACAG	<p>CC CG CU- AA-- UU</p> <p>CUCUG UCC UGC ACUGAGCUG ACACAG \</p> <p>GGGAC AGG ACG UGACUCGGU UGUGUU G</p> <p>A- -- ACU CACA UG</p>
miR-25	CAUUGCACUUGUCUGGGUCUGA	<p>A AG G UU G UG ACG</p> <p>GGCC GUGUUG AGGC GAGAC G GCAAU CUGG C</p> <p>CCGG CGUGAC UCUG CUCUG C CGUUA GGUC U</p> <p>C AG G UU A CG CCG</p>
miR-26a	UUCAGUAUAUCCAGGAUAGGCU	<p>AGCC GUG CCUCGU CAAGUAA CCAGGAUAGGCUU GCAG</p> <p>UCCGG CGC GGGGCA GUUCAUU GGUUCUAUCCGGUA U</p> <p>G A C - ACCC</p>
miR-26b	UUCAGUAUAUCCAGGAUAGGCU	<p>GA - U UC UGUG</p> <p>CCGG CCC AGU CAAGUAAU AGGAUAGGUUG \</p> <p>GGCC GGG UCG GUUCAUUA UCUGUCCGAC C</p> <p>AG C - CC CUGU</p>
miR-27a	UUCACAGUGGCUAAGUUCGCGU	<p>A A A U G UCCAC</p> <p>CUG GG GC GGGCUUAGCUGCU GUGAGCA GG \</p> <p>GAC CC CG CUUGAAUCCGGUA CACUUGU CU A</p> <p>C C C - G GAACC</p>



Fig. 7 (cont.)

miR-27b	UUCACAGUGGCUAAAGUUCUG	AGGUGCAGAGCUUAGCUG    AUUG    UGAU    U UCCACGUCUUGAAUCGGU    GUAACAG    UGG \ GA--    UC--    U
miR-28	AAGGAGCUCACAGUCUAUUGAG	C    A    U    ----    CC GGU CUUGCCCUCC AGGAGCUCACAGUCUA UG    AGUUA U UCA GGACGGGAG UCCUCGAGUGUUAGAU AC    UCAGU U G    C    CCUU    CU
miR-29a	CUAGCACCAUCUGAAAUUCGGUU	UUU    C    UCAAU AUGACUGAUUUC    UGGUGUU AGAG \ UAUUGGCUAAAG    ACCACGA UCUU    A UCU    -    UUAUU
miR-29b	UAGCACCAUUUGAAAUUCAGUGUU	A    U    GU    UUAUU AGGA GCUGGUUUA AUGGUG UUAAGU \ UCUU UGACUAAAGU UACCAC GAUCUG    A G    U    --    UUAGUG
miR-29c	UAGCACCAUUUGAAAUUCGGUua	
miR-30a-s	UGUAAACAUCUCCUGGACUGGAAGC	A    UC    ----    A GCG CUGUAAACAUCUCC GACUGGAAGCU    GUG A CGU GACGUUUUGUAGG CUGACUUUCGG    CAC G --    GUAAA    C
miR-30a-as	CUUUCAGUCGGAUGUUUUCAGC	A    UC    ----    A GCG CUGUAAACAUCUCC GACUGGAAGCU    GUG A CGU GACGUUUUGUAGG CUGACUUUCGG    CAC G C    --    GUAGA    C

Fig. 7 (cont.)

miR-30b	UGUAAACAUCUACACUCAGC	<p>U - UCAUA  <u>AUGUAAACAUC</u> <u>ACA</u> <u>CUCAGCUG</u> C  UGCAUUUGUAGG UGU GGGUCGGU A  - A UGCGU</p>
miR-30c	UGUAAACAUCUACACUCUCAGC	<p>UACU U ACA GUGGAA  AGA GUAAACA CCU CUCUCAGCU A  UCU CAUUUGU GGA GAGGGUCGA G  UUCU C A-- AAGAAU human</p>
miR-30d	UGUAAACAUCUACACUCUGGAG	<p>U U CCC GUAAGA  GU GU GUAAACAUC GACUGGAAGCU C  CA CG CGUUUGUAG CUGACUUUCGA A  U U A-- AUCGAC chr8 human</p>
miR-31	GGCAAGAUGCUGGCAUAGCUG	<p>GA G C U- GAA  GGAGAG GGCAA AUG UGGCAUAGC GUU C  CCUUUC CCGUU UAC ACCGUAUCG CAA U  UA A A UC GGG</p>
miR-32	UAUUGCACAUUACUAGUUGC	<p>U - UU C  GGAGAUAUUGCACAU ACUAGUUGCAU G GU A  CUUUUAUAGUGUGUG UGAUUUAACGUA C CG C  - A UC G</p>
miR-33	GUGCAUUGUAGUUGCAUUG	<p>A UU UUCU UG  CUGUGGUGCAUUGU G GCAUUGCAUG GG  GACACUACGUGACA C UGUACGUAC CC G  C UU ---- AU</p>
miR-99a	ACCCGUAGAUCGGAUCUUGU	<p>A UC U G AAG  CAUA ACCCGUAGA CGA CUUGUG UG U  GUGU UGGGUAUCU GCU GAACGC GC G  C UU C - CAG</p>

Fig. 7 (cont.)

miR-99b	CACCCGUAGAACCGACCUUGCG	<p>CC AC C ----- C</p> <p>GGCAC ACCCGUAGA CGA CU UGCGG GG \</p> <p>CUGUG UGGGUGUCU GCU GA ACGCC CU C</p> <p>CC GU C ACAC G U</p>
miR-101	UACAGUACUGUGAUACUGA	<p>A GUCCA U</p> <p>UCAGUUAUACACAGUCUG UGCU U</p> <p>AGUCAUAGUGUCAUGAC AUGG U</p> <p>- AAAUC</p>
miR-122a	UGGAGUGUGACAAUGGUGUUUGU	<p>GG C UGUCC</p> <p>AGCUGU AGUGUGA AAUGGUGUUUG A</p> <p>UCGAUA UCACACU UUACCGCAAC A</p> <p>AA A UAUA</p> <p>woodchuck</p>
miR-122b	UGGAGUGUGACAAUGGUGUUUGA	
miR-122a,b	UGGAGUGUGACAAUGGUGUUUG	
miR-123	CAUUAUUACUUUUGGUACGCG	<p>A A U CGCUG C</p> <p>UGAC GC CAUUAUUACUU UGUACG UGA A</p> <p>ACUG CG GUAAUUAUGAG GCCAUGC ACU C</p> <p>G C U UCAA- U</p>
miR-124a*	UUAAGGCACGCGGUGAUGCCA	<p>- C A GA UAAUG</p> <p>CUCU G GUGUUCAC GCG CCUUGAUU U</p> <p>GAGA C CGUAAGUG CGC GGAUUA C</p> <p>A - G AC CAUUA</p>

Fig. 7 (cont.)

miR-124b	UUAAGGCACGCGGGUGAAUGC	CC A GA UAAUG CUCU GUGUUCAC GCG CCUUGAUU GAGA CGUAAGUG CGC GGAUUAU U AC G AC CAUAC AC021518
miR-125a	UCCUGAGACCCUUUAACCUUG potential lin-4 ortholog	C UA CUGGUU CCUGAGA CCUU ACCUGUGA GG C GGUCCG GGUUUCU GGAG UGGACACU CC G A U -- GGGA U
miR-125b	UCCUGAGACCCUUAACCUUGUGA potential lin-4 ortholog	UC C A GG- U GCCUAG CCUGAGA CCU ACUUGUGA UAU U CGGAUC GGUUUCU GGA UGAACACU AUG U CA U C ACA A
miR-126	UCGUACCGUGAGUAAUAAUGC	A U CGCUG C GC CAUUAUUAUUU UGGUACG UGA A CG GUAAUUAUAG GCCAUGC ACU C C U UCAA- U
miR-127	UCGGAUCCGUCUGAGCUUGGCU	A U G G C -- AG CC GCC GCU AAGCUCAGA GG UCUGAU UC GG UGG CGG UUCGAGUCU CC AGGCUA AG A C U - G U CU AA
miR-128	UCACAGUGAACCGGUCUCUUUU	UUC UAG CU U GUUGGA GGGGCCG CACUGU GAGAGGU U CGACUU CUCUGGC GUGACA CUCUUUA A UUU CAA -- C
miR-129	CUUUUUUCGGUCUGGGCUUGC	- C CU G UUCU C GGAU CUUUUUUG GGU GGGCUU CUG CU A UCUA GAAAAAC CCA CCCGAA GAC GA A U C UU G UGAU- C human

Fig. 7 (cont.)

miR-130	CAGUGCAAUGUUAAAAGGGC	<p>- C A GUCUAAAC GA GCUCUUUU ACAUUGUGCU CU CU CGGGA AAA UGUAACGUGA GA G A U C G CCAUGU</p>
miR-131	UAAAGCUAGAUAAACCGAAAGU	<p>G C G U A GUU UUAU UUUGGUUAUCUAGCU UAUGAG GU U CAA AAUG AAGCCAAUAGAUGCA AUACUU UG U A A A C G</p>
miR-132	UAAACAGUCUACAGCCAUGGUCGU	<p>A UUC G- G GGGC ACCGUGGCU GAUUGUUACU UGG CCCG UGGUACCGA CUGACAAUGG GCC A C CAU AG A</p>
miR-133	UUGGUCCCCUUAACCAAGCUGU	<p>A AA U A GCCUC GCUA AGCUGGU AA GG ACCAAUUC U CGAU UCGACCA UU CC UGGUUUAG U G AC C C GUAAC</p>
miR-134	UGUGACUGGUUGACCAGAGGGA	<p>GU U A- G GCGU AC AGGGU GUGACUGG UG CCA AGGG GC UCCCA CACUGAUC AC GGU UCCC UG U AC C CG G ACU- UC</p>
miR-135	UAUGGCUUUUUUAUCCUAUGUGAA	<p>UUU UUCUUAU CUAUGGCUUU AUUCCUAUGUGA GGUGCCGAGG UAGGGAUAUACU U U- CGCUCG</p>
miR-136	ACUCCAUUUGUUUUGAUGAUGGA	<p>C UUU UUCU GAGGACUC AUUUG UGAUGAUGGA CUUCUGAG UAAAC GCUACUACCU U - UCU CGAA</p>

Fig. 7 (cont.)

miR-137	UAUUGC UUAAGAAUACGCGUAG	<p>CUUCGGU ACG GUUUCUUGGGUGG UAAUA CG \</p> <p>GGAGCUG UGC CAUAAAGAAUUCGUU AUUGU GC U</p> <p>A G - - U AU</p>
miR-138	AGCUGGUGUUGUGAAUC	<p>-- UCA AC- C CG</p> <p>CAGCU GGUGUUGUGAA GGCCG GAG AG C</p> <p>GUUGG CCACAGCACUU UCGGC UUC UC A</p> <p>GA UA- CCA - CU</p>
miR-139	UCUACAGUGCAGGUGUCU	<p>G - U A GUGGC</p> <p>GU UAUUCUA CAG GC CGUGUCUCCAGU \</p> <p>CA AUGAGGU GUC CG GCGCAGAGGUCG U</p> <p>U C - GAGGC human</p>
miR-140	AGUGGUUUUACCCUUAUGGUAG	<p>- A A UU UC</p> <p>CCUG CC GUGGUUUUACCCU UGGUAGG ACG A</p> <p>GGAC GG CACCAAGAUGGA ACCAUCU UGU U</p> <p>A - - CG</p>
miR-141	AACACUGUCUGGUAAAGAUGG	<p>U -- U AU GAAG</p> <p>GGG CCAUCUU CCAG GCAGUGUUGG GGUU \</p> <p>CCC GGUAGAA AU - C- AGUA</p>
miR-142s	CAUAAAGUAGAAAGCACUAC	<p>AC- A UAA--- G</p> <p>CCAUAAAGUAG AAGCACUAC CA C</p> <p>GGUAAUUCauc UUUGUGAUG GU A</p> <p>GUA C UGGGAG C</p>
miR-142as*	UGUAGUGUUUCCUACUUUAUGG	<p>AC- A UAA--- G</p> <p>CCAUAAAGUAG AAGCACUAC CA C</p> <p>GGUAAUUCauc UUUGUGAUG GU A</p> <p>GUA C UGGGAG C</p>

Fig. 7 (cont.)

new	AUAAGACGAGCAAAAGCUUGU	<p>G C C G G C AU  UGAC GCGAGCUUU GC CG UUAUAC UG \</p> <p>ACUG UGUUCGAAA CG GC AAUAUG AC G</p> <p>G A A AG C UC</p> <p>AL049829.4</p>
miR-143	UGAGAUGAAGCACUGAGCUca UUAGAUGAAGCACUGUAG	<p>G G U - AG  CCUGAG UGCAGUCU CAUCUC GG UC U</p> <p>GGACUC AUGUCACGA GUAGAG CU AG U</p> <p>G A U G GG</p> <p>AC008681.7</p>
miR-144	UACAGUAUAGAUGAUGUACUAG	<p>G A A- GU  GGCUGG AUAUCAUC UAUACUGUA GUUU G</p> <p>CUGAUC UGUAGUAG AUAUGACAU CAGA A</p> <p>A - CA GU</p>
miR-145	GUCCAGUUUUCCCAGGAUCCCUU	<p>C UC U C UGGAUG  CUCA GG CAGU UU CCAGGAUCCCU \</p> <p>GAGU UC GUCA AA GGUCCUAGGGG C</p> <p>- UU U A UAGAAU</p>
miR-146	UGAGAACUGAAUCCAUUGGGUUU	<p>CU C AUAUC  AGCU GAGAACUGAAUU CAUGGGUU A</p> <p>UCGA UUCUUGACUUA GUGUCCAG A</p> <p>C- A ACUGU</p>
miR-147	GUGUGUGGAAAUAGCUUCUGCC	<p>A- CAA ACA--- GA  AAUCUA AGA CAUUUCUGCACAC CCA \</p> <p>UUAGAU UCU GUAAAGGUGUGUG GGU C</p> <p>CG UC- ACCGAA AU human</p>
miR-148	UCAGUGGCACUACAGAAUUUUGU	<p>- A- CC - AGU  GAGGCAAGUUCUG AG CACU GACU CUG \</p> <p>CUCUGUUUCAAGAC UC GUGA CUGA GAU A</p> <p>A AC -- A AGU human</p>

Fig. 7 (cont.)

miR-149	UCUGGCUCGUGUCUUCACUCC	<p> <u>G C G A</u> GUG G  GGCUCUG CUC GU UCUUC CUCCC UUU U  UCGGGC GAG CA GGAGG GAGG GAG C  G A G - AG- C </p>
miR-150	UCUCCCAACCCUUGUACCAGUGU	<p> AC U UG- UG  CCUGUCUCCCA CCU GUACCAG CUG \  GGGAUAGGGGU GGA CAUGGUC GAC C  CC - CCA UC </p>
miR-151	CUAGACUGAGGCUCUUGAGGU	<p> C CA UGUCU  CCUG CCUCGAGGAGCU CAGUCUAGUA \  GGAC GGAGUCCUCGG GUCAGAUCAU C  A A- CCCUC </p>
miR-152	UCAGUGCAUGACAGAACUUGG	<p> G A CC CGG C  CCGGCCUAGGUUCUGU AU CACU GACU GCU U  GGCCCGGUUCAGACA UA GUGA CUGA CGA G  G C - - - - G </p>
miR-153	UUGCAUAGUCACAAAAGUGA	<p> GU A- AAU  CAGUG UCAUUUUUGUGAU UGCAGCU GU \  GUUAC AGUAAAAACACUG ACUUUGA CG A  U AU CC AGU </p>
miR-154	UAGGUUAUCCGUGUUGCCUUCG	<p> U - CCU-- UUU  GAAGAUAAGGUA CCGUGU UG UCGC \  UUUUUAUCCAGU GGCACA AC AGUG A  U U UAAGC UUU </p>
miR-155 [BIC-RNA]	UUAAUGCUAAUUGUGAUAGGGG	<p> U U A UUGGCC  CUGUUAUAGCUAAU G G UAGGGGU \  GACAAUUAACGAUG U C AUCCUCAG U  - C - UCAGUC </p>



Fig. 7 (cont.)

name	sequence	structure
miR-C1	AACAUUCACGCUGCGGUGAGU	<pre> U  A  U  CU  A  GGAUUA CCA GG ACA UCAACG GUCGGUG GUU GGU CC UGU AGUUGC CAGCCAC CAA  A U  A  C  --  -  AAAACAAA </pre>
miR-C2	UUUGGCAUUGGUAGAACUCACA	<pre> UU  UGG  UCA  UAAGGU ACCAU  UUGGCAA  UAGAAC  CACCGG  A UGGUA  AACCGUU  AUCUUG  GUGGCC  A UC      CAG      ---  CAGGGU </pre>
miR-C3	UAUGGCACUGGUAGAAUUCACUG	<pre> G  AC--  GA  --  AC CUGU UAUGGC  UGGUA  AUUCACUG  UGA  A GACA AUACCG  GCCAU  UAAGUGAC  ACU  G A      GGAA  --  UG  CU </pre>
miR-C4	CUUUUUGCGGUCUGGGCUUGUU	<pre> -  C  CU  G  UUUU  C UGGAU CUUUUG  GUU  GGGCUU  CUG  CU  G AUCUA GAAAAAC  CCA  CCCGAA  GAC  GA  A U      C  UU  G  UGAU  C </pre>
miR-C5	UGGACGGAGAACUGAUAAAGGGU	<pre> U      C  AG  -  UG CCU  UCCUUAUA  UUUUCC  CCAGC  UUUG  A GGA  GGGAAUAGU  AAGAGG  GGUUG  GAAU  C U      C  CA  U  CU </pre>
miR-C6	UGGAGAGAAAGGCAGUUC	<pre> A  G  AU  UC AGGGAUUGGAG  GAAAG  CAGUUCUG  GG  C UUCCUGGUCUC  CUUUC  GUGGGGAC  CC  C -      G  --  UC </pre>

Fig 7 (cont.)

name	sequence	structure
miR-C7	CAAAGAAUUCUCCUUUUUGGGCUU	<pre>       U  UU  UCUCAU  U ACUUUCCAAAGAAUUC  CCUU  GGGCUU  U UGAAGGGUUUUUUAAG  GGAA  CCCGAA  U       U  U-  UUUUAU </pre>
miR-C8	UCGUGUCUUGUGUUGCAGCCGG	<pre>       A  A  C  CGCUGC UC  GGCU  CAACACACAGGAC  CGGG  U GG  CCGA  GUUGUGUUCUG  GCUC  C       -  C  U  CCCAGU </pre>
miR-C9	UAAACACUGUCUGGUAACGAUGU	<pre>       C  UU  UUG GGGCAUC  UUACCGGACAGUG  UGGA  UC  \ CUGGUAG  AAUGGUCUGUCAC  AUCU  AG  G       C  A  C-  UUC </pre>
miR-C10	CAUCCCUUGCAUGGUGGAGGGU	<pre>       CA  UC  GU  UGAGCUC UCU  CA  CCUUGCAUG  GGAGGG  U AGG  GU  GGGACGUAC  CCUCCC  C       AC  UU  AC  CAAAAGU </pre>
miR-C11	GUGCCUACUGAGCUGACAUCAGU	<pre>       G  G  A  UA  UCUCAU CUCC  GU  CCU  CUGAGCUGA  UCAGU  \ GAGG  CA  GGA  GACUUGACU  GGUCA  U       A  A  C  C-  CACACU </pre>
miR-C12	UGAUAUGUUUGAUAUAUUAGGU	<pre>       U-  UA---  UU CUGUG  GAUAUGUUUGAUAUAU  GGUUG  \ GACAU  UUAUACGAACUAUAUA  CUAAU  A       CC  UCAAC  UU </pre>

Fig. 7 (cont.)

name	sequence	structure
miR-C13	CAACGGAAUCCCAAAAGCAGCU	<pre> C      C  AA  UU  -  C AGCGG AACGAAUCC AA GCAGCUG GU CU C UCGUCC UUGCUUAGG UU CGUCGAC UA GA A C      -  CA  CU  C  G </pre>
miR-C14	CUGACCUAUGAAUUGACA	<pre> C      -  A  UGCUCUC UGACCUAUG AAUUG CAGCCAG G ACUGGAUAC UUAAC GUCGGUC U -      C  UCCCCUC </pre>
miR-C15	UACCACAGGGUAGAACCACGGA	<pre> -  G      A  UU  UC UCCUG CCG UGGUUUACCCU UGGUAGG ACG A AGGAC GGC ACCAAGAUGGGA ACCAUCU UGU U A      -  C  --  CG </pre>
miR-C16	AACUGGCCUACAAGUCCCGAG	<pre> A      U  C  A  A  AGU GAG GCUGGG CUUUG GGGC AG UGAG G CUC UGACCC GAAAC UCCG UC ACUU U C      U  A  G  A  GAC </pre>
miR-C17	UGUAAACAGCAACUCCAUGUGGA	<pre> U      A  G  -  -  U AUCGGG GUAACAGCA CUCCAU UGGA CUG G UAGUCU CADUGUCGU GAGGUG ACCU GGC C U      C  -  UA  U </pre>
miR-C18	UAGCAGCACAGAAAUUUGGC	<pre> U      A-  UG  GAA AGCAGCACAG AAUAUUGGCA GG G UCGUCGUGUC UUAUAACCGU CU U GG  --  GAG </pre>

Fig 7 (cont.)

name	sequence	structure
miR-C19	UAGGUAGUUUCAUGUUGUUGG	<pre> A  A  C  GGCCUGGG GUGAAUU GGU GUUU AUGUUUGUUG  U CACUUAG CCA CAAA UACAACAAC  U C  C  C  U  ACAAGUCU </pre>
miR-C20	UUCACCACCUUCUCCACCCAGC	<pre> C  A  CA  GA  -  A GGCUGUGC GGU GAGAGGG GUGG GGU AAG G CCGGUACG CCCA CUCUUC CACU CCA UUC C A  C  AC  UC  C  U </pre>
miR-C21	GGUCCAGAGGGGAGAUAGG	<pre> G  -  C  G  U  UUCUUG UCAUU G UC A AGGGAGA AGG  U AGUAA U AG U UCUCUUCU UCC  G A  A  A  A  -  UUUUUA </pre>
miR-C22	CCCAGUGUUCAGACUACCUGUU	<pre> AAC  U  C  U  G--- G GCC CCAGUGU CAGACUAC UGU CA  GAG \ CGG GGUUACA GUCUGAUG ACA GU  CUC C AUU  C  -  U  GUAA  U </pre>
miR-C23	UAAUACUGCCUGGUAUAUGAUGAC	<pre> GGC  -  C  UAGUG GCCGU CAUC UUACUGGGCAG AUUGGA  U CGGCA GUAG AAUGGUCCGUC UAAUCU  C ---  U  A  CUAGU </pre>
miR-C24	UACUCAGUAAGGCAUUGUUCU	<pre> U  U  UUC  A UACCUUAC CAG AAGGCAUUGUUC  UAU U AUGGGAUG GUC UCCCGUGACAAG  AUA U U  U  UAA  A </pre>

Fig.7 (cont.)

name	sequence	structure
miR-C25	AGAGGUAUAGCGCAUGGGAAGA	<pre>       U      A-      UG      C GUUCC UUUUCCUADGC UAUACUUCUU UGGAU \ CGAGG AGAAGGGUACG AUAUGGAGAA AUCUG U       U      CG      --      G </pre>
miR-C26	UGAAAUUUUUAGGACCACUAG	<pre>       C      U      G      A      C      U GGUC AGUGGUUCU GACA UUCA CAGUU UG \ CCAG UCACCAGGA UUGU AAGU GUUAA AC A       A      U      A      -      C      G </pre>
miR-C27	UUCCCUUUUGUCAUCCUAUGGCCUG	<pre>       U      A      U      GAGAAUA UGGAC UCCCUUUGUC UCCUA GCCU \ ACUUG AGGGAACGG AGGGU CGGA U       C      A      -      GGAAGUA </pre>
miR-C28	UCCUUCAUUCCACCGGAGUCUG	<pre>       UC      C      UCUUA CUCUUG CUUCAUCCAC GGAGUCUG U GAGGAC GAAGUGAGGUG CUUAGAC G       UC      A      CAACC </pre>
miR-C29	GUGAAAUUUUAGGACCACUAGA	<pre>       U      C      U      G      A      C      U GCC GGUC AGUGGUUCU GACA UUCA CAGUU UG \ CGG CCAG UCACCAGGA UUGU AAGU GUUAA AC A       C      A      U      A      -      C      G </pre>
miR-C30	UGGAAUGUAAGGAGUGUGUGG	<pre>       -      C      U      AUAUC CCAGG CCACAUGCUUUUAU C CAUAG \ GGUUU GGUGUGAAGGAUGUA G GUAUC U       U      A      -      ACGAC </pre>

Fig 7 (cont.)

name	sequence	structure
miR-C31	UACAGUAGUCUGCACAUUGGUU	<pre> AUC      U      C      ----- G GCC      CCAGUGU CAGACUAC UGU      UCAG A CGG      GGUUACA GUCUGAUG ACA      GGUC G AUU      C      -      UGUACAG      G </pre>
miR-C32	CCCUGUAGAACCGAAUUUUGUGU a miR-10 variant	<pre> A      G      C      UG- AC UAUAU CCCU UAGAA CGAAUUUUGU GU C AUAUA GGGG AUUUU GCUUAGACAC UA C A      -      A      UGA CA </pre>
miR-C33	AACCCGUAGAUCCGAACUUGUGA A a miR-99a variant	<pre> A      C      C      A      C      AU CACA ACC GUAGAU CGA CUUGUG UG U GUGU UGG UAUCUG GUU GAACAC AC C A      A      U      C      -      GU </pre>
miR-C34	GCUUCUCCUGGCUCUCCUCCUCCUC	<pre> C      U      UUG      GGAG AAGG AGGGG GAGGGG CGGGAGGAGC CGGGC G UCC UCUCU CUCCUC GUCCUCCUCCG GUUCG C -      -      UCG      C      GCGU </pre>

Fig. 7 (cont)

name	human	C. elegans	mouse							Drosophila	fugu fish	zebrafish
			liver	small intes	colon	cerebellum	cortex	midbrain	heart	spleen		
let-7a-1	AC007924 chr9 AC087784 chr 17 identical precursor		num. hits in trace data, 3 families of similar precursors		found		nearly identical precursor	found				
let-7a-2	AP001359 chr11						nearly identical precursor					
let-7a-3	AL049853 chr22	AP274345 chrX with diff. precursor								AB001659 diff. precursor		
let-7b	AL049853 chr22		nearly identical precursor			nearly ident precursor tracel48311003		found	EST AF481799.1 spleen cerebellum (mammary)		with slightly diff precursor	
let-7c	AP001667 chr21		identical and diff. precursors			num. genomic hits, ident precursor; diff precursor -> EST AI614897	numerous genomic hits	found				
let-7d	AC007924.3 chr9 AC087784 chr17 identical			found		tracel83587042 nearly ident prec	tracel83587042 2 nearly ident prec	found	found			
let-7e	AC018755 chr19							found	FOUND			
let-7f-1	AC007924 chr9 AC007704 chr17					ident precursor genomic DNA		found	found			
let-7f-2	AI592046 chrX					ident. precursor in smtrace 18713911						
let-7g	precursor ident. to mouse in AC092045.2 chr3					genomic hits, no EST		found				
let-7h							found in cortex, no db hit					

let-7f	precursor ident. to mouse [AL117383.19]; SLSO AC018341.22				found, supported by EST AB661268	found						
mir-1											2L, AB003667	
mir-1b	AL449263.5 chr20 nt1-21	U97403.1 nt 1-21 (22G)	no mouse hit (only nt1-21)			found	found					
mir-1c								found, but no db hit				
mir-1d	AL449263.5 chr20 nt1-22 (23G)					found	trace hits (nt1- 23) trace#91 523974				BF157601.1 with C23 (diff. precursor)	
mir-2a-1										2L, AB003663		
mir-2a-2										2L, AB003663		
mir-2b-1										2L, AB003620		
mir-2b-2										2L, AB003663		
mir-3										2R, AB003795		
mir-4										2R, AB003795		



Fig. 7 (cont.)

miR-5																		2R, AE003795				
miR-6-1																		2R, AE003795				
miR-6-2																		2R, AE00379				
miR-6-3																		2R, AE00379				
miR-7	AC003791 chr19 diff.precursor; EST AP171191 again different																	not cloned, but mouse EST predicts precursor similar to human				
miR-8																		2R, AE003605				
miR-9	AC003116 chr15 AC026701 chr5 each with diff. precursor																	3L, AE003516	2diff precurs scaffold- 3868 and 2417			
miR-10	AP287967 chr11 (H0X B4/B5)																	AE001574				
miR-11																		not found, but AC01194 chr.11 predicts diff. precursor				
miR-12																		3R, AE003735				
miR-13a																		X, AE003490				
																		3R, AE003708				

Fig. 7 (cont.)

[illegible]

[illegible]

Fig. 7 (cont.)

mir-27a	19, AC020916				found	found, but no db hit		found, but no db hit for mouse	found	found		
mir-27b	XK_09843.1 chr9 identical precursor							found, maps to chr 13 MGSC mmtrace 44671617				
mir-28	3, AC063932											
mir-29a	7, AP017104 second ident.copy found in chr7 CLUSTER, this cluster also consvd in mouse!				found, AC024913.32	found, mmtrace#23467334		nearly ident precursor trace#2346733 4, EST AC024913.32	found	found, EST, nearly ident prec		
mir-29b	AL035209.1 chr1 CLUSTER of mir-29-b and 29-c; miRNA similar to mir-83				found			AC024913.32, d iff precursor in EST B0342396 (retina)	found	found	Scaffold 17670. (A third copy)	
mir-29c								found	found	found, supported by ESTs	Scaffold 17670 has two copies of this RNA	
mir-30a-s	nearly ident fold in AL035467.23 chr6						found, ESTs, trace#802 3889 all with 22G	found	found			
mir-30a-s	6, AL035467							found with diff. precursor in trace #8261735				
mir-30b	human AF159227.6 chr8, different precursor							trace#72329251	found	found	Scaffold 3483, diff precursor	
mir-30c	AL136164.0 chr.6 supported by ESTs (BF594736.1)							found, but no db hit for mouse		found		

Fig. 7 (cont.)

[illegible]

Fig. 7 (cont.)

miR-124a*	nearly ident. precursor in chr8(AC021518) chr20(AL096828) precursor	found in 272504.1 chr1V intron, diff precursor	found	most abundant in celeb., genomic hits (trace#1097008, 11737241) found, but no db hit	most abundant;seve ral trace hits;precurs- orabellum	found	slightly diff precursor AC009251 chr2L		
miR-124b	AC021518 chr8, nearly ident. chr20 AL096828.29			hit					
miR-125a	ident. precu in AC018755.3 chr 19			genomic hits trace#33921945, 48262259 and more	found				
miR-125b	AP001359.4 chr11 AP001667.1 chr21(chr21 like mouse)				trace#8398570 5	found with A22U	found in AC006590.1 1 with diff fold	scaffold_ 2358	
miR-126				matrace#3521597 and more		found		with diff precurs scaffold_32 95	
miR-127	human AL117190.6 chr.14 same precurs as in mouse			hit in trace#79514537					
miR-128	ident in AC016742.10 chr 21 diff prec in AC016943.7 chr.3			genomic hit trace#51670230	found	found		scaffold_ 828, diff- prec	
miR-129	human AC016662.3 chr7			found, but no db hit					
miR-130				mmtrace 68479278					with diff fold AC091299.2
miR-131	AC005317.2 chr 15 sligh. diff precursor, but AC026701.6 chr 5 ident			several trace hits, mouse AP155142	found				
miR-132	AL137038.5 chr17 prec sligh. diff from mouse			trace hit#86984641					

Fig. 7 (cont.)

miR-133	AL391221.15 chr6 diff. precursor (ident to rat L23722.1)						found, trace1 62107955	found	AC093440.1 diff. precursor	Scaffold_ 1049; prec u nearly like mouse	
miR-134	AL132709.5 chr14 similar precursor						trace16462031 1				
miR-135	AC092045.2 chr3 AC018659.35 chr12 (ident or simil to mouse)						trace17149323 5, EST 85989780995 1(kidn, apc en) (-chr3 huma n)	found		Scaffold_ 2125 with similar precurs	
miR-136	AL117190.6 chr14 ident to mouse						trace18607175 3				
miR-137	AC027691.1 chr1 , ident to mouse, nearly ident fish						trace18977454 3, EST (hypothal) AR 52436.1, ident .			Scaffold_ 18294 nearly ident to mouse/han	
miR-138	AC006058.1 chr3 precursor diff						mouse EST BB528620.2				
miR-139	AP003065.2 chr11						found, but no mouse hit				
miR-140	AC026468.8 chr.16, precursor r nearly ident,					several trace hits) trace1053 0393					
miR-141	AC006512.12 chr12, precursor slightly diff					AC007397 chr6		found			
miR-142a	AC004687.1 chr17 BCL1/myq translocation locus, like mouse					found					
miR-142as*						several EST A115235		found			

Fig. 7 (cont.)

[illegible]





Fig. 7 (cont.)

name	human	mouse					Drosophila	fugu fish	zebrafish
		spleen	eye	kidney	testes	lung			
mir-C1	with different precursors in chr9 AL158075.11, chr1 AL136321.5		mouse trace #78647842			found		scaffold_1819	
mir-C2	chr7 AC084864.2 similar precursor		mouse trace #88841093					scaffold_967	AL590150.2
mir-C3	chr7 AC084864.2 ident. precursor		trace #86029980					scaffold_967	AL590150.2
mir-C4	similar precurs. in chr7 AC018662.3		trace #13885686		found				
mir-C5	chr15 AC069082.9		trace #87318220				found	scaffold_3671	
mir-C6	chr22 AC005664.2 ident. precursor		chr16 AC012526.32						
mir-C7	chr1 AL512443.7 similar prec.		trace #86694995						
mir-C8				found, trace #51673384					
mir-C9				found, trace #78964803				scaffold_2210, diff. precursor	
mir-C10	chrX AP222686.1 nearly ident. precursor			found, trace #61928192					
mir-C11	chr9 XM_098943.1 has C170; prec. nearly identical to mouse			found, cDNA AL286629.1, has C170					
mir-C12				found, trace#71760450				scaffold_2294	
mir-C13		found		found, trace #88722637					

Fig. 7 (cont.)

name	human	mouse						Drosophila	fugu fish	zebrafish
		spleen	eye	kidney	testes	lung	thymus			
mir-C14	chr11 AC000159.6			found, but no db hit						
mir-C15	chr16 AC026468.6 nearly ident.precursor			EST B1687377.1, several trace					scaffold_2003	
mir-C16	chr17 AC003101.1, similar precursor			found, trace#95 55103					scaffold_246	
mir-C17	chr11 AC000159.6, chr1 AC103590.2; diff.prec.			found, trace #87796602					scaffold_152	
mir-C18				found, trace #47823768 (close to mir-16)		found		found		
mir-C19	chr17 AC009789.21 cloned from human cell line only			similar precursor in mouse chr11 AC011194.15					scaffold_18334	
mir-C20	chr1 AL355310.19 cloned from human cell line only									
mir-C21	chr3 AC063952.15 cloned from human cell line only									
mir-C22	chr19 AC007229.1; chr1 AL137157.7 similar precursor; cloned from human cell line only								scaffold_8399	
mir-C23					trace #72257777	found			scaffold_2210	
mir-C24					trace #69879879					
mir-C25					trace #49754566					
mir-C26	AL136001 ident. precursor				trace #11977216					

Fig. 7 (cont.)

name	human	mouse							Drosophila	fugu fish	zebrafish
		spleen	eye	kidney	testes	lung	thymus	skin			
mir-C27	chr9 AL159990.12 identical precursor		trace #91503159							scaffold_725	
mir-C28	XM 036612.4. precursor very similar							XN_149012.1		scaffold_13664	
mir-C29	chr14 AL136001.6 nearly identical precursor							trace #18453604			
mir-C30	chr6 AL391221.15 similar precursor							trace #84055510			
mir-C31	chr9 AC006312.8							trace #89079710		scaffold_5830	
mir-C32								U77364.1, intronic location Hoxd4 gene		scaffold_82	
mir-C33								trace #84780544		scaffold_15612	
mir-C34							trace# 72109322				

## SEQUENCE LISTING

<110> Max-Planck-Gesellschaft zur Förderung der Wissensc

<120> Identification of novel genes encoding for small  
temporal RNA

<130> 26250PWO\_DR

<140>

<141>

<150> EP 01 123 453.1

<151> 2001-09-28

<150> EP 02 006 712.0

<151> 2002-03-22

<150> EP 02 016 772.2

<151> 2002-07-26

<160> 217

<170> PatentIn Ver. 2.1

<210> 1

<211> 29

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:  
Oligonucleotide

<400> 1

tactatacaa cctactacct caatttgcc

29

<210> 2

<211> 21

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:  
Oligonucleotide

<400> 2

actatgcaac ctactacctc t

21

<210> 3

<211> 21

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:  
Oligonucleotide

<400> 3

actatacaac ctctacctc a

21

<210> 4

<211> 19

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:  
Oligonucleotide

<400> 4

tggtgtttcc gcccgga

19

<210> 5

<211> 22

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:  
Oligonucleotide

<400> 5

tggaatgtaa agaagtatgg ag

22

<210> 6

<211> 23

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:  
Oligonucleotide

<400> 6

gctcctcaaa gctggctgtg ata

23

<210> 7

<211> 22

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:  
Oligonucleotide

<400> 7

tgagacacac tttgccagtg ga

22

<210> 8

<211> 21

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:  
Oligonucleotide

<400> 8

tcaatggttg tctagcttta t

21

<210> 9

<211> 23

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:  
Oligonucleotide

<400> 9

catatcacao cgatcggtcc ttt

23

<210> 10

<211> 22

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:  
Oligonucleotide

<400> 10

aaaaagaaca gccactgtga ta

22

<210> 11

<211> 23

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:  
Oligonucleotide

<400> 11

tggaagacta gtgattttgt tgt

23

<210> 12

<211> 23

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:  
Oligonucleotide

<400> 12

gacatcttta cctgacagta tta

23

<210> 13

<211> 23

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:  
Oligonucleotide

<400> 13

tcatacagct agataaccaa aga

23



<210> 14  
<211> 21  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence:  
Oligonucleotide

<400> 14  
acaaattcgg atctacaggg t 21

<210> 15  
<211> 21  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence:  
Oligonucleotide

<400> 15  
gcaagaactc agactgtgat g 21

<210> 16  
<211> 23  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence:  
Oligonucleotide

<400> 16  
accagtacct gatgtaatac tca 23

<210> 17  
<211> 22  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence:

## Oligonucleotide

&lt;400&gt; 17

actcgtcaaa atggctgtga ta

22

&lt;210&gt; 18

&lt;211&gt; 21

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

<223> Description of Artificial Sequence:  
Oligonucleotide

&lt;400&gt; 18

taggagagag aaaaagactg a

21

&lt;210&gt; 19

&lt;211&gt; 21

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

<223> Description of Artificial Sequence:  
Oligonucleotide

&lt;400&gt; 19

tagcagcaca taatggtttg t

21

&lt;210&gt; 20

&lt;211&gt; 21

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

<223> Description of Artificial Sequence:  
Oligonucleotide

&lt;400&gt; 20

gccaatatattt acgtgctgct a

21

&lt;210&gt; 21

&lt;211&gt; 22

&lt;212&gt; DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:  
Oligonucleotide

<400> 21

tacaagtgcc ttactgcag ta

22

<210> 22

<211> 22

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:  
Oligonucleotide

<400> 22

tatctgcact agatgcacct ta

22

<210> 23

<211> 23

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:  
Oligonucleotide

<400> 23

tcagttttgc atagatttgc aca

23

<210> 24

<211> 22

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:  
Oligonucleotide

<400> 24

tacctgcact ataagcactt ta

22

<210> 25  
<211> 22  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence:  
Oligonucleotide

<400> 25  
tcaacatcag tctgataagc ta 22

<210> 26  
<211> 22  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence:  
Oligonucleotide

<400> 26  
acagttcttc aactggcagc tt 22

<210> 27  
<211> 21  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence:  
Oligonucleotide

<400> 27  
ggaaatccct ggcaatgtga t 21

<210> 28  
<211> 22  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence:  
Oligonucleotide

<400> 28

ctgttcctgc tgaactgagc ca

22

<210> 29

<211> 22

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:

Oligonucleotide

<400> 29

tcagaccgag acaagtgcaa tg

22

<210> 30

<211> 22

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:

Oligonucleotide

<400> 30

agcctatcct ggattacttg aa

22

<210> 31

<211> 22

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:

Oligonucleotide

<400> 31

agcggaactt agccactgtg aa

22

<210> 32

<211> 22

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:  
Oligonucleotide

<400> 32

ctcaatagac tgtgagctcc tt

22

<210> 33

<211> 22

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:  
Oligonucleotide

<400> 33

aaccgatttc agatggtgct ag

22

<210> 34

<211> 22

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:  
Oligonucleotide

<400> 34

gctgcaaaca tccgactgaa ag

22

<210> 35

<211> 22

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:  
Oligonucleotide

<400> 35

cagctatgcc agcatcttgc ct

22

<210> 36  
<211> 21  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence:  
Oligonucleotide

<400> 36  
gcaacttagt aatgtgcaat a 21

<210> 37  
<211> 22  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence:  
Oligonucleotide

<400> 37  
tgcaatgcaa ctacaatgca cc 22

<210> 38  
<211> 22  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence:  
Oligonucleotide

<400> 38  
ctccatactt ctttacattc ca 22

<210> 39  
<211> 21  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence:  
Oligonucleotide

<400> 39  
gctgagtgtta ggatgtttac a 21

<210> 40  
<211> 23  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence:  
Oligonucleotide

<400> 40  
gcttccagtc gaggatgttt aca 23

<210> 41  
<211> 22  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence:  
Oligonucleotide

<400> 41  
cgcaagggtcg gttctacggg tg 22

<210> 42  
<211> 20  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence:  
Oligonucleotide

<400> 42  
tcagttatca cagtactgta 20

<210> 43  
<211> 23  
<212> DNA  
<213> Artificial Sequence



<220>

<223> Description of Artificial Sequence:  
Oligonucleotide

<400> 43

acaaacacca ttgtcacact cca

23

<210> 44

<211> 21

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:  
Oligonucleotide

<400> 44

tggcattcac cgcggtgcctt a

21

<210> 45

<211> 23

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:  
Oligonucleotide

<400> 45

cacaggttaa aggggtctcag gga

23

<210> 46

<211> 22

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:  
Oligonucleotide

<400> 46

tcacaagtta ggggtctcagg ga

22

<210> 47

<211> 22  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence:  
Oligonucleotide

<400> 47  
agccaagctc agacggatcc ga

22

<210> 48  
<211> 22  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence:  
Oligonucleotide

<400> 48  
aaaagagacc gggtcactct ga

22

<210> 49  
<211> 21  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence:  
Oligonucleotide

<400> 49  
gcaagcccag accgaaaaaa g

21

<210> 50  
<211> 20  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence:  
Oligonucleotide

<400> 50

gcccttttaa cattgcactc

20

<210> 51

<211> 21

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:  
Oligonucleotide

<400> 51

actttcgggtt atctagcttt a

21

<210> 52

<211> 23

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:  
Oligonucleotide

<400> 52

acgaccatgg ctgtagactg tta

23

<210> 53

<211> 22

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Combined DNA/RNA Molecule:  
Oligonucleotide

<220>

<223> Description of Artificial Sequence:  
Oligonucleotide

<400> 53

tgagctacag tgcttcactc ca

22

<210> 54

<211> 18

<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence:  
Oligonucleotide

<400> 54  
uuuaaccgcg aattccag 18

<210> 55  
<211> 18  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence:  
Oligonucleotide

<400> 55  
acggaattcc tcactaaa 18

<210> 56  
<211> 23  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence:  
Oligonucleotide

<400> 56  
cacaggtaa aggtctcag gga 23

<210> 57  
<211> 24  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence:  
Oligonucleotide

<400> 57  
cagccaacg aattcctcac taaa 24

<210> 58  
<211> 22  
<212> RNA  
<213> D. melanogaster

<400> 58  
uggaauguaa agaaguaugg ag 22

<210> 59  
<211> 23  
<212> RNA  
<213> D. melanogaster

<400> 59  
uaucaacagcc agcuuugaug agc 23

<210> 60  
<211> 23  
<212> RNA  
<213> D. melanogaster

<400> 60  
uaucaacagcc agcuuugagg agc 23

<210> 61  
<211> 22  
<212> RNA  
<213> D. melanogaster

<400> 61  
ucacugggca aagugugucu ca 22

<210> 62  
<211> 21  
<212> RNA  
<213> D. melanogaster

<400> 62  
auaaagcuag acaaccuug a 21

<210> 63

<211> 23  
<212> RNA  
<213> D. melanogaster

<400> 63  
aaaggaacga ucguugugau aug 23

<210> 64  
<211> 22  
<212> RNA  
<213> D. melanogaster

<400> 64  
uaucaacagug gcuguucuuu uu 22

<210> 65  
<211> 23  
<212> RNA  
<213> D. melanogaster

<400> 65  
uggaagacua gugauuuugu ugu 23

<210> 66  
<211> 23  
<212> RNA  
<213> D. melanogaster

<400> 66  
uaauacuguc agguaaaagau guc 23

<210> 67  
<211> 23  
<212> RNA  
<213> D. melanogaster

<400> 67  
ucuuugguua ucuagcugua uga 23

<210> 68  
<211> 21  
<212> RNA  
<213> D. melanogaster

&lt;400&gt; 68

accuguaga uccgaauug u

21

&lt;210&gt; 69

&lt;211&gt; 21

&lt;212&gt; RNA

&lt;213&gt; D. melanogaster

&lt;400&gt; 69

caucacaguc ugaguucug c

21

&lt;210&gt; 70

&lt;211&gt; 23

&lt;212&gt; RNA

&lt;213&gt; D. melanogaster

&lt;400&gt; 70

ugaguauuac aucagguacu ggu

23

&lt;210&gt; 71

&lt;211&gt; 22

&lt;212&gt; RNA

&lt;213&gt; D. melanogaster

&lt;400&gt; 71

uauacacagcc auuuugacga gu

22

&lt;210&gt; 72

&lt;211&gt; 22

&lt;212&gt; RNA

&lt;213&gt; D. melanogaster

&lt;400&gt; 72

uauacacagcc auuuugauga gu

22

&lt;210&gt; 73

&lt;211&gt; 21

&lt;212&gt; RNA

&lt;213&gt; D. melanogaster

&lt;400&gt; 73

ucagucuuuu ucucucuccu a

21

<210> 74  
<211> 22  
<212> RNA  
<213> D. melanogaster

<400> 74  
ugagguagua gguuguauag uu 22

<210> 75  
<211> 22  
<212> RNA  
<213> Human

<400> 75  
ugagguagua gguuguauag uu 22

<210> 76  
<211> 22  
<212> RNA  
<213> Human

<400> 76  
ugagguagua gguugugugg uu 22

<210> 77  
<211> 22  
<212> RNA  
<213> Human

<400> 77  
ugagguagua gguuguaugg uu 22

<210> 78  
<211> 21  
<212> RNA  
<213> Human

<400> 78  
agagguagua gguugcauag u 21

<210> 79



<211> 21  
<212> RNA  
<213> Human

<400> 79  
ugagguagga gguuguauag u

21

<210> 80  
<211> 22  
<212> RNA  
<213> Human

<400> 80  
ugagguagua gauuguauag uu

22

<210> 81  
<211> 22  
<212> RNA  
<213> Human

<400> 81  
uagcagcaca uaaugguuug ug

22

<210> 82  
<211> 22  
<212> RNA  
<213> Human

<400> 82  
uagcagcacg uaaauauugg cg

22

<210> 83  
<211> 20  
<212> RNA  
<213> Human

<400> 83  
acugcaguga aggcacuugu

20

<210> 84  
<211> 22  
<212> RNA  
<213> Human

&lt;400&gt; 84

uaaggugcau cuagugcaga ua

22

&lt;210&gt; 85

&lt;211&gt; 23

&lt;212&gt; RNA

&lt;213&gt; Human

&lt;400&gt; 85

ugugcaaauc uaugcaaaac uga

23

&lt;210&gt; 86

&lt;211&gt; 23

&lt;212&gt; RNA

&lt;213&gt; Human

&lt;400&gt; 86

ugugcaaauc caugcaaaac uga

23

&lt;210&gt; 87

&lt;211&gt; 22

&lt;212&gt; RNA

&lt;213&gt; Human

&lt;400&gt; 87

uaaagugcuu auagugcagg ua

22

&lt;210&gt; 88

&lt;211&gt; 22

&lt;212&gt; RNA

&lt;213&gt; Human

&lt;400&gt; 88

uagcuauca gacugauguu ga

22

&lt;210&gt; 89

&lt;211&gt; 22

&lt;212&gt; RNA

&lt;213&gt; Human

&lt;400&gt; 89

aagcugccag uugaagaacu gu

22

<210> 90  
<211> 21  
<212> RNA  
<213> Human

<400> 90  
aucacauugc cagggauuuc c 21

<210> 91  
<211> 22  
<212> RNA  
<213> Human

<400> 91  
uggcucaguu cagcaggaac ag 22

<210> 92  
<211> 22  
<212> RNA  
<213> Human

<400> 92  
cauugcacuu gucucggucu ga 22

<210> 93  
<211> 22  
<212> RNA  
<213> Human

<400> 93  
uucaaguaau ccaggauagg cu 22

<210> 94  
<211> 22  
<212> RNA  
<213> Human

<400> 94  
uucaaguaau ucaggauagg uu 22

<210> 95

<211> 22

<212> RNA

<213> Human

<400> 95

uucacagugg cuaaguuccg cu

22

<210> 96

<211> 22

<212> RNA

<213> Human

<400> 96

aaggagcuca cagucuauug ag

22

<210> 97

<211> 22

<212> RNA

<213> Human

<400> 97

cuagcaccau cugaaaucgg uu

22

<210> 98

<211> 22

<212> RNA

<213> Human

<400> 98

cuuucagucg gauguuugca gc

22

<210> 99

<211> 21

<212> RNA

<213> Human

<400> 99

ggcaagaugc uggcauagcu g

21

<210> 100

<211> 21

<212> RNA

<213> Human

<400> 100  
uauugcacau uacuaaguug c 21

<210> 101  
<211> 19  
<212> RNA  
<213> Human

<400> 101  
gugcauugua guugcauug 19

<210> 102  
<211> 22  
<212> RNA  
<213> Human

<400> 102  
uggaauguaa agaaguaugg ag 22

<210> 103  
<211> 23  
<212> RNA  
<213> Human

<400> 103  
uggaagacua gugauuuugu ugu 23

<210> 104  
<211> 23  
<212> RNA  
<213> Human

<400> 104  
uccuugguua ucuagcugua uga 23

<210> 105  
<211> 21  
<212> RNA  
<213> Human

<400> 105  
accugugaga uccgaauuug u 21

<210> 106  
<211> 22  
<212> RNA  
<213> Mouse

<400> 106  
ugagguagua gguuguauag uu

22

<210> 107  
<211> 22  
<212> RNA  
<213> Mouse

<400> 107  
ugagguagua gguugugugg uu

22

<210> 108  
<211> 22  
<212> RNA  
<213> Mouse

<400> 108  
ugagguagua gguuguaugg uu

22

<210> 109  
<211> 21  
<212> RNA  
<213> Mouse

<400> 109  
agagguagua gguugcauag u

21

<210> 110  
<211> 21  
<212> RNA  
<213> Mouse

<400> 110  
ugagguagga gguuguauag u

21

<210> 111

<211> 22  
<212> RNA  
<213> Mouse

<400> 111  
ugagguagua gauuguauag uu

22

<210> 112  
<211> 22  
<212> RNA  
<213> Mouse

<400> 112  
ugagguagua guuuguacag ua

22

<210> 113  
<211> 22  
<212> RNA  
<213> Mouse

<400> 113  
ugagguagua guguguacag uu

22

<210> 114  
<211> 19  
<212> RNA  
<213> Mouse

<400> 114  
ugagguagua guuugugcu

19

<210> 115  
<211> 22  
<212> RNA  
<213> Mouse

<400> 115  
uggaauaguaa agaaguaugu aa

22

<210> 116  
<211> 22  
<212> RNA  
<213> Mouse

&lt;400&gt; 116

uggaauguaa agaaguaugu ac

22

&lt;210&gt; 117

&lt;211&gt; 23

&lt;212&gt; RNA

&lt;213&gt; Mouse

&lt;400&gt; 117

uggaauguaa agaaguaugu auu

23

&lt;210&gt; 118

&lt;211&gt; 23

&lt;212&gt; RNA

&lt;213&gt; Mouse

&lt;400&gt; 118

ucuuugguua ucuagcugua uga

23

&lt;210&gt; 119

&lt;211&gt; 22

&lt;212&gt; RNA

&lt;213&gt; Mouse

&lt;400&gt; 119

uagcagcaca uaaugguuug ug

22

&lt;210&gt; 120

&lt;211&gt; 22

&lt;212&gt; RNA

&lt;213&gt; Mouse

&lt;400&gt; 120

uagcagcaca ucaugguuua ca

22

&lt;210&gt; 121

&lt;211&gt; 22

&lt;212&gt; RNA

&lt;213&gt; Mouse

&lt;400&gt; 121

uagcagcacg uaaauauugg cg

22



<210> 122

<211> 22

<212> RNA

<213> Mouse

<400> 122

uaaggugcau cuagugcaga ua

22

<210> 123

<211> 23

<212> RNA

<213> Mouse

<400> 123

ugugcaaauc caugcaaaac uga

23

<210> 124

<211> 23

<212> RNA

<213> Mouse

<400> 124

uaaagugcuu auagugcagg uag

23

<210> 125

<211> 22

<212> RNA

<213> Mouse

<400> 125

uagcuauca gacugaugu ga

22

<210> 126

<211> 22

<212> RNA

<213> Mouse

<400> 126

aagcugccag uugaagaacu gu

22

<210> 127

<211> 21  
<212> RNA  
<213> Mouse

<400> 127  
aucacauugc cagggauuuc c 21

<210> 128  
<211> 23  
<212> RNA  
<213> Mouse

<400> 128  
aucacauugc cagggauuac cac 23

<210> 129  
<211> 22  
<212> RNA  
<213> Mouse

<400> 129  
uggcucaguu cagcaggaac ag 22

<210> 130  
<211> 22  
<212> RNA  
<213> Mouse

<400> 130  
uucaaguaau ccaggauagg cu 22

<210> 131  
<211> 22  
<212> RNA  
<213> Mouse

<400> 131  
uucaaguaau ucaggauagg uu 22

<210> 132  
<211> 22  
<212> RNA  
<213> Mouse

<400> 132  
uucacagugg cuaaguuccg cu 22

<210> 133  
<211> 20  
<212> RNA  
<213> Mouse

<400> 133  
uucacagugg cuaaguucug 20

<210> 134  
<211> 22  
<212> RNA  
<213> Mouse

<400> 134  
cuagcaccau cugaaaucgg uu 22

<210> 135  
<211> 23  
<212> RNA  
<213> Mouse

<400> 135  
uagcaccauu ugaaaucagu guu 23

<210> 136  
<211> 22  
<212> RNA  
<213> Mouse

<400> 136  
uagcaccauu ugaaaucggu ua 22

<210> 137  
<211> 23  
<212> RNA  
<213> Mouse

<400> 137  
uguaaacauc cucgacugga agc 23

<210> 138  
<211> 22  
<212> RNA  
<213> Mouse

<400> 138  
cuuucagucg gauguuugca gc 22

<210> 139  
<211> 21  
<212> RNA  
<213> Mouse

<400> 139  
uguuaaacauc cuacacucag c 21

<210> 140  
<211> 23  
<212> RNA  
<213> Mouse

<400> 140  
uguuaaacauc cuacacucuc agc 23

<210> 141  
<211> 22  
<212> RNA  
<213> Mouse

<400> 141  
uguuaaacauc cccgacugga ag 22

<210> 142  
<211> 20  
<212> RNA  
<213> Mouse

<400> 142  
acccguagau ccgaucuugu 20

<210> 143

<211> 22

<212> RNA

<213> Mouse

<400> 143

cacccguaga accgaccuug cg

22

<210> 144

<211> 20

<212> RNA

<213> Mouse

<400> 144

uacaguacug ugauaacuga

20

<210> 145

<211> 23

<212> RNA

<213> Mouse

<400> 145

uggaguguga caaugguguu ugu

23

<210> 146

<211> 23

<212> RNA

<213> Mouse

<400> 146

uggaguguga caaugguguu uga

23

<210> 147

<211> 22

<212> RNA

<213> Mouse

<400> 147

uggaguguga caaugguguu ug

22

<210> 148

<211> 21

<212> RNA

<213> Mouse

&lt;400&gt; 148

cauuauuacu uuugguacgc g

21

&lt;210&gt; 149

&lt;211&gt; 22

&lt;212&gt; RNA

&lt;213&gt; Mouse

&lt;400&gt; 149

uuaaggcacg cggugaaugc ca

22

&lt;210&gt; 150

&lt;211&gt; 21

&lt;212&gt; RNA

&lt;213&gt; Mouse

&lt;400&gt; 150

uuaaggcacg cgggugaaug c

21

&lt;210&gt; 151

&lt;211&gt; 23

&lt;212&gt; RNA

&lt;213&gt; Mouse

&lt;400&gt; 151

ucccugagac ccuuuaaccu gug

23

&lt;210&gt; 152

&lt;211&gt; 22

&lt;212&gt; RNA

&lt;213&gt; Mouse

&lt;400&gt; 152

ucccugagac ccuaacuugu ga

22

&lt;210&gt; 153

&lt;211&gt; 21

&lt;212&gt; RNA

&lt;213&gt; Mouse

&lt;400&gt; 153

ucguaccgug aguaauaauug c

21

<210> 154  
<211> 22  
<212> RNA  
<213> Mouse

<400> 154  
ucggauccgu cugagcuugg cu 22

<210> 155  
<211> 22  
<212> RNA  
<213> Mouse

<400> 155  
ucacagugaa ccggucucuu uu 22

<210> 156  
<211> 21  
<212> RNA  
<213> Mouse

<400> 156  
cuuuuuucgg ucugggcuug c 21

<210> 157  
<211> 20  
<212> RNA  
<213> Mouse

<400> 157  
cagugcaaug uaaaaagggc 20

<210> 158  
<211> 21  
<212> RNA  
<213> Mouse

<400> 158  
uaaagcuaga uaaccgaaag u 21

<210> 159

<211> 23

<212> RNA

<213> Mouse

<400> 159

uaacagucua cagccauggu cgu

23

<210> 160

<211> 22

<212> RNA

<213> Mouse

<400> 160

uugguccccu ucaaccagcu gu

22

<210> 161

<211> 22

<212> RNA

<213> Mouse

<400> 161

ugugacuggu ugaccagagg ga

22

<210> 162

<211> 24

<212> RNA

<213> Mouse

<400> 162

uauggcuuuu uauuccuaug ugaa

24

<210> 163

<211> 23

<212> RNA

<213> Mouse

<400> 163

acuccauuug uuuugaugau gga

23

<210> 164

<211> 22

<212> RNA

<213> Mouse



&lt;400&gt; 164

uauugcuuua gaauacgcu ag

22

&lt;210&gt; 165

&lt;211&gt; 17

&lt;212&gt; RNA

&lt;213&gt; Mouse

&lt;400&gt; 165

agcugguguu gugaauc

17

&lt;210&gt; 166

&lt;211&gt; 18

&lt;212&gt; RNA

&lt;213&gt; Mouse

&lt;400&gt; 166

ucuacagugc acgugucu

18

&lt;210&gt; 167

&lt;211&gt; 21

&lt;212&gt; RNA

&lt;213&gt; Mouse

&lt;400&gt; 167

agugguuuua ccuauaggua g

21

&lt;210&gt; 168

&lt;211&gt; 21

&lt;212&gt; RNA

&lt;213&gt; Mouse

&lt;400&gt; 168

aacacugucu gguaaagau g

21

&lt;210&gt; 169

&lt;211&gt; 20

&lt;212&gt; RNA

&lt;213&gt; Mouse

&lt;400&gt; 169

cauaaaguag aaagcacuac

20

<210> 170  
<211> 22  
<212> RNA  
<213> Mouse

<400> 170  
uguaguguuu ccuacuuuau gg 22

<210> 171  
<211> 22  
<212> RNA  
<213> Mouse

<400> 171  
ugagaugaag cacuguagcu ca 22

<210> 172  
<211> 22  
<212> RNA  
<213> Mouse

<400> 172  
uacaguauag augauguacu ag 22

<210> 173  
<211> 24  
<212> RNA  
<213> Mouse

<400> 173  
guccaguuuu cccaggauc ccuu 24

<210> 174  
<211> 23  
<212> RNA  
<213> Mouse

<400> 174  
ugagaacuga auuccauggg uuu 23

<210> 175

<211> 21  
<212> RNA  
<213> Mouse

<400> 175  
guguguggaa augcuucugc c 21

<210> 176  
<211> 22  
<212> RNA  
<213> Mouse

<400> 176  
ucagugcacu acagaacuuu gu 22

<210> 177  
<211> 22  
<212> RNA  
<213> Mouse

<400> 177  
ucuggcuccg ugucuucacu cc 22

<210> 178  
<211> 23  
<212> RNA  
<213> Mouse

<400> 178  
ucucccaacc cuuguaccag ugu 23

<210> 179  
<211> 22  
<212> RNA  
<213> Mouse

<400> 179  
cuagacugag gcuccuugag gu 22

<210> 180  
<211> 21  
<212> RNA  
<213> Mouse

<400> 180  
ucagugcaug acagaacuug g 21

<210> 181  
<211> 20  
<212> RNA  
<213> Mouse

<400> 181  
uugcauaguc acaaaaguga 20

<210> 182  
<211> 22  
<212> RNA  
<213> Mouse

<400> 182  
uagguuaucc gugugccuu cg 22

<210> 183  
<211> 22  
<212> RNA  
<213> Mouse

<400> 183  
uuaaugcuaa uugugauagg gg 22

<210> 184  
<211> 23  
<212> RNA  
<213> Human/Mouse

<400> 184  
aacauucaac gcugucggug agu 23

<210> 185  
<211> 22  
<212> RNA  
<213> Human/Mouse

<400> 185  
uuuggcaaug guagaacuca ca 22

<210> 186

<211> 23

<212> RNA

<213> Human/Mouse

<400> 186

uauggcacug guagaauuca cug

23

<210> 187

<211> 22

<212> RNA

<213> Human/Mouse

<400> 187

cuuuuugcgg ucugggcuug uu

22

<210> 188

<211> 22

<212> RNA

<213> Human/Mouse

<400> 188

uggacggaga acugauaagg gu

22

<210> 189

<211> 18

<212> RNA

<213> Human/Mouse

<400> 189

uggagagaaa ggcaguuc

18

<210> 190

<211> 23

<212> RNA

<213> Human/Mouse

<400> 190

caaagaauuc uccuuuuggg cuu

23

<210> 191

<211> 22

<212> RNA

<213> Human/Mouse

<400> 191

ucgugucuug uguugcagcc gg

22

<210> 192

<211> 21

<212> RNA

<213> Human/Mouse

<400> 192

uaacacuguc ugguaacgau g

21

<210> 193

<211> 22

<212> RNA

<213> Human/Mouse

<400> 193

caucccuugc augguggagg gu

22

<210> 194

<211> 23

<212> RNA

<213> Human/Mouse

<400> 194

gugccuacug agcugacauc agu

23

<210> 195

<211> 22

<212> RNA

<213> Human/Mouse

<400> 195

ugauauguuu gauauuuag gu

22

<210> 196

<211> 22

<212> RNA

<213> Human/Mouse

&lt;400&gt; 196

caacggauc ccaaaagcag cu

22

&lt;210&gt; 197

&lt;211&gt; 18

&lt;212&gt; RNA

&lt;213&gt; Human/Mouse

&lt;400&gt; 197

cugaccuaug aaugaca

18

&lt;210&gt; 198

&lt;211&gt; 22

&lt;212&gt; RNA

&lt;213&gt; Human/Mouse

&lt;400&gt; 198

uaccacaggg uagaaccacg ga

22

&lt;210&gt; 199

&lt;211&gt; 21

&lt;212&gt; RNA

&lt;213&gt; Human/Mouse

&lt;400&gt; 199

aacuggccua caaaguccca g

21

&lt;210&gt; 200

&lt;211&gt; 22

&lt;212&gt; RNA

&lt;213&gt; Human/Mouse

&lt;400&gt; 200

uguaacagca acuccaugug ga

22

&lt;210&gt; 201

&lt;211&gt; 21

&lt;212&gt; RNA

&lt;213&gt; Human/Mouse

&lt;400&gt; 201

uagcagcaca gaaauauugg c

21

<210> 202

<211> 21

<212> RNA

<213> Human/Mouse

<400> 202

uagguaguuu cauguuguug g

21

<210> 203

<211> 22

<212> RNA

<213> Human/Mouse

<400> 203

uucaccaccu ucuccaccca gc

22

<210> 204

<211> 19

<212> RNA

<213> Human/Mouse

<400> 204

gguccagagg ggagauagg

19

<210> 205

<211> 22

<212> RNA

<213> Human/Mouse

<400> 205

cccaguguuc agacuaccug uu

22

<210> 206

<211> 23

<212> RNA

<213> Human/Mouse

<400> 206

uaauacugcc ugguaaugau gac

23

<210> 207



<211> 21

<212> RNA

<213> Human/Mouse

<400> 207

uacucaguaa ggcauguuc u

21

<210> 208

<211> 22

<212> RNA

<213> Human/Mouse

<400> 208

agagguauag cgcaugggaa ga

22

<210> 209

<211> 21

<212> RNA

<213> Human/Mouse

<400> 209

ugaauguuu aggaccacua g

21

<210> 210

<211> 23

<212> RNA

<213> Human/Mouse

<400> 210

uucccuugu cauccaugc cug

23

<210> 211

<211> 22

<212> RNA

<213> Human/Mouse

<400> 211

uccucauuc caccggaguc ug

22

<210> 212

<211> 23

<212> RNA

<213> Human/Mouse

<400> 212  
gugaaauguu uaggaccacu aga

23

<210> 213  
<211> 22  
<212> RNA  
<213> Human/Mouse

<400> 213  
uggaauguaa ggaagugugu gg

22

<210> 214  
<211> 22  
<212> RNA  
<213> Human/Mouse

<400> 214  
uacaguaguc ugcacauugg uu

22

<210> 215  
<211> 22  
<212> RNA  
<213> Human/Mouse

<400> 215  
cccuguagaa ccgaauuugu gu

22

<210> 216  
<211> 24  
<212> RNA  
<213> Human/Mouse

<400> 216  
aaccgcuaga uccgaacuug ugaa

24

<210> 217  
<211> 23  
<212> RNA  
<213> Human/Mouse

<400> 217  
gcuucuccug gcucuccucc cuc

23

**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

**BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☐ **FADED TEXT OR DRAWING**
- ☐ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☐ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☐ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER:** \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**